

## SEARCH REQUEST FORM

Requestor's Name: Patricia Delfy Serial Number: 68/938,548  
Date: 18 Aug Phone: 305-7555 Art Unit: 1645  
Rev 8205

### Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

Please search SEQJ DND: 2  
4  
6  
8  
9  
10  
12

SEARCHED  
SERIALIZED  
INDEXED  
FILED

Please include an interference search.

Thanks  
Pat.

2

Amendment Due.

### STAFF USE ONLY

Date completed: 8/21/94  
Searcher: Schepman  
Terminal time: \_\_\_\_\_  
Elapsed time: \_\_\_\_\_  
CPU time: \_\_\_\_\_  
Total time: \_\_\_\_\_  
Number of Searches: \_\_\_\_\_  
Number of Databases: \_\_\_\_\_

Search Site  
\_\_\_\_ STIC  
\_\_\_\_ CM-1  
\_\_\_\_ Pre-S  
Type of Search  
\_\_\_\_ N.A. Sequence  
\_\_\_\_ A.A. Sequence  
\_\_\_\_ Structure  
\_\_\_\_ Bibliographic

Vendors  
\_\_\_\_ IG  
\_\_\_\_ STN  
\_\_\_\_ Dialog  
\_\_\_\_ APS  
\_\_\_\_ Geninfo  
\_\_\_\_ SDC  
\_\_\_\_ DARC/Questel  
\_\_\_\_ Other

\*\*\*\*\*

W P E R E H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 20:53:06 1999; MasPar time 12.09 Seconds  
591.342 Million cell updates/sec

Tabular output not generated.

Title: >US-08-938-548B-2  
Description: (1-131) from US08938548B.pep  
Perfect Score: 931  
Sequence: 1 MNLPTKVSWAATVLLLL.....GRRCSAPAAASVAPGGQSGI 131

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl9

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phage 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp Vertebrate 14:sp\_virus

Statistics: Mean 39.956; Variance 91.033; scale 0.439

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	931	100.0	131	4	043612	PREPRO-OREXIN.	1.03e-143
2	845	90.8	131	6	077668	PREPRO-OREXIN PRECURSOR	5.14e-128
3	782	84.0	130	11	055232	PREPRO-OREXIN	1.46e-116
4	775	83.2	130	11	055241	HYPOCRETIN (PREPRO-ORE	2.72e-115
5	110	11.8	1321	4	075129	KIAA0634 PROTEIN (FRAG	2.12e-02
6	103	11.1	679	10	023352	HYPOTHETICAL 74.6 KD P	1.60e-01
7	101	10.8	1584	4	014514	BAI 1.	2.81e-01
8	100	10.7	480	4	060455	POLYADENYLATE BINDING	3.71e-01
9	96	10.3	205	13	042394	NISHED (FRAGMENT)	1.12e+00
10	96	10.3	400	4	060609	GDNF FAMILY RECEPTOR A	1.12e+00
11	96	10.3	453	10	049019	MYB-LIKE DNA-BINDING D	1.12e+00
12	95	10.2	226	2	063119	HYPOTHETICAL 23.1 KD P	1.47e+00
13	95	10.2	382	2	087918	SODIUM DEPENDENT PHOSP	1.93e+00
14	94	10.1	1337	4	015255	PROTEIN-TYROSINE PHOSP	1.93e+00
15	94	10.1	1736	4	075111	KIAA0612 PROTEIN (FRAG	1.93e+00
16	93	10.0	469	9	037839	ORF469 PROTEIN.	2.52e+00
17	92	9.8	833	4	099523	SORTILIN PRECURSOR.	3.29e+00
18	91	9.8	175	4	060895	MRNA ENCODING RAMP2 PR	4.29e+00
19	90	9.7	147	11	061639	3' ORF.	5.58e+00
20	90	9.7	574	11	035298	ACYLOXYACYL HYDROLASE.	5.58e+00

21 90 9.7 901 4 060391 R32184\_2 5.58e+00  
22 90 9.7 990 10 040699 LEUCINE-RICH REPEAT/RE 5.58e+00  
23 89 9.6 172 11 061104 FLT3 LIGAND, T169 FORM 7.24e+00  
24 89 9.6 4848 2 007944 PRISTINAMYCIN I SYNTHA 9.38e+00  
25 88 9.5 480 4 092743 NOVEL SERINE PROTEASE. 9.38e+00  
26 88 9.5 519 3 013420 PHENOLOXIDASE (EC 1.10 9.38e+00  
27 88 9.5 1174 6 095168 TIGHT JUNCTION PROTEIN 9.38e+00  
28 87 9.3 84 11 060471 ANION EXCHANGER ISOFOR 1.21e+01  
29 87 9.3 551 2 072405 PCBR. 1.21e+01  
30 87 9.3 729 11 060470 ANION EXCHANGER 2 A (F 1.21e+01  
31 87 9.3 1534 6 082898 RIBOSOME RECEPTOR. 1.57e+01  
32 86 9.2 266 2 007405 MAV266 (FRAGMENT). 1.57e+01  
33 86 9.2 277 2 033285 HYPOTHETICAL 30.9 KD P 1.57e+01  
34 86 9.2 920 2 087342 BETA-(1-3)-GLUCOSYL TR 1.57e+01  
35 86 9.2 2195 3 002822 VESICLE COAT PROTEIN S 1.57e+01  
36 85 9.1 100 10 043535 ORF PRECURSOR (FRAGMEN 2.02e+01  
37 85 9.1 139 2 005844 VRF53 PROTEIN (FRAGMEN 2.02e+01  
38 85 9.1 143 4 015412 CTGA. 2.02e+01  
39 85 9.1 331 13 091640 LEUCINE ZIPPER WITH BA 2.02e+01  
40 85 9.1 397 10 048758 F21B7.28. 2.02e+01  
41 85 9.1 422 4 007111 GLIAL GROWTH FACTOR 2 2.02e+01  
42 85 9.1 760 11 062178 SEMAPHORIN B PRECURSOR 2.02e+01  
43 85 9.1 767 14 066627 ORF 24. 2.02e+01  
44 85 9.1 804 5 044896 ZK484.4 PROTEIN. 2.02e+01  
45 85 9.1 974 4 043306 KIAA0422 (FRAGMENT). 2.02e+01

#### ALIGNMENTS

RESULT 1  
ID 043612 PRELIMINARY; PRT; 131 AA.  
AC 043612;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE PREPRO-OREXIN.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98150861.  
RA SAKURAI T., ANEMIYA A., ISHII M., MATSUZAKI I., CHEMELLI R.M.,  
RA TANAKA H., WILLIAMS S.C., RICHARDSON J.A., KOZLOWSKI G.P., WILSON S.,  
RA ARCH J.R.S., BUCKINGHAM R.E., HAYNES A.C., CARR S.A., ANNAN R.S., D.J.,  
RA MCNULTY D.E., LIU W.-S., TERRETT J.A., ELSHOURBAGY N.A., BERGMA D.J.,  
RA YANAGISAWA M.;  
RT "Orexins and orexin receptors: a family of hypothalamic neuropeptides  
and G protein-coupled receptors that regulate feeding behavior.";  
RL CELL 92:573-585(1998).  
DR EMBL; AF041240; G2897118; -;  
SQ SEQUENCE 131 AA; 13363 MW; 2C11048A CRC32;

Query Match 100.0%; Score 931; DB 4; Length 131;  
Best Local Similarity 100.0%; Pred. No. 1.03e-143;  
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 MNLPTKVSWAATVLLLLPPALLSGAAAPLPCCRCQKTCSCRLYELHAGNHA 60  
|||||  
QY 1 MNLPTKVSWAATVLLLLPPALLSGAAAPLPCCRCQKTCSCRLYELHAGNHA 60  
|||||  
Db 61 AGITLGRKRRSGPGLQRLQRLQASGNHAGILTMGRAGAEPAAPRCLGRRCSAPAA 120  
|||||  
QY 61 AGITLGRKRRSGPGLQRLQRLQASGNHAGILTMGRAGAEPAAPRCLGRRCSAPAA 120  
|||||  
Db 121 ASVAPGGQSGI 131  
|||||  
QY 121 ASVAPGGQSGI 131  
|||||

RESULT 2  
ID 077668 PRELIMINARY; PRT; 131 AA.  
AC 077668;

```

Matches 107; Conservative 13; Mismatches 10; Indels 1; Gaps 1;

Db 1 MNFSTKVPRAVTLTLLLL-PPALLSLGVDAQPLDCCRCQKTCSCRLYELLHGAGNHA 59
QY 1 MNLPSTVSWAAVTLTLLLL-PPALLSSGAAQPLDCCRCQKTCSCRLYELLHGAGNHA 60

Db 60 AGILTLGKRGGPGLOGRLQRLQNGNHAAGILTMGRRAGALEPHPCSGRCPTVTT 119
QY 61 AGILTLGKRGGPGLOGRLQRLQNGNHAAGILTMGRRAGAPAPRCLGLRCSAPAA 120

Db 120 TALAPRGSGV 130
QY 121 ASVAPGQSGI 131

RESULT 5
ID 075129 PRELIMINARY; PRT; 1321 AA.
AC 075129;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE KIAA0634 PROTEIN (FRAGMENT).
GN KIAA0634.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RC TISSUE=BRIN;
RX MEDLINE; 98403880.
RA ISHIKAWA K., NAGASE T., SUYAMA M., MIYAJIMA N., TANAKA A., KOTANI H.,
RA NOMURA N., OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.;"
RL DNA RES. 5:169-176(1998).
DR EMBL; AB014534; D1032570; -.
FT NON_TER 1
SQ SEQUENCE 1321 AA; 145424 MW; 4B1721D3 CRC32;

Query Match 11.8%; Score 110; DB 4; Length 1321;
Best Local Similarity 57.6%; Pred. No. 2.12e-02;
Matches 19; Conservative 7; Mismatches 4; Indels 3; Gaps 3;

Db 68 LLLFLLLPPPLLAGATAAASREPDSPCRCLKT 100
QY 15 LLLLLLLPPA-LLSSGAAQPL-PDC-CROKT 44

RESULT 6
ID 023352 PRELIMINARY; PRT; 679 AA.
AC 023352;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 74.6 KD PROTEIN.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRSS).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RC SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
OS SEQUENCE FROM N.A.
RX MEDLINE; 98121113.
RA BEVAN M., BANCROFT I., BENT E., LOVE K., GOODMAN H., DEAN C.,
RA BERGAMP R., DIRKSE W., VAN STAVEREN M., STIEKEMA W., DROST L.,
RA RIDLEY P., HUDSON S.A., PATEL K., MURPHY G., PIFFANELLI P., WEDLER H.,
RA WEDLER E., WAMBUIT R., WEITZENEGER T., POHL T.M., TERRY N.,
RA GLEEN J., VILLARROEL R., DE CLERCK R., VAN MONTAGU M., LECHARNY A.,
RA AUBORG S., GY I., KREIS M., LAO N., KAVANAGH T., HEMPEL S., KOTTER P.,

```

```

RA ENTIAN K.D., RIEGER M., SCHAEFFER M., FUNK B., MUELLER-AUER S.,
RA SILVEY M., JAMES R., MONTFORT A., PONS A., PUIGDOMENECH P., DOUKA A.,
RA VOUKEATOU E., MILTONI D., HATZOPOULOS P., PIRAVANDI E., OBERMAIER B.,
RA HILBERT H., DUESTERHOFT A., MOORES T., JONES J.D.G., ENEVA T.,
RA PALME K., BENES V., RECHMAN S., ANSORGE W., COOKE R., BERGER C.,
RA DELSENY M., VOET M., VOLCKAERT G., MEWES H.W., KLOSTERMAN S.,
RA SCHUELLER C., CHALWATZIS N.;
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana."
RL NATURE 391:485-488(1998).
DR EMBL; Z97337; E326841; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 679 AA; 74635 MW; B301B713 CRC32;

Query Match 11.1%; Score 103; DB 10; Length 679;
Best Local Similarity 30.6%; Pred. No. 1.60e-01;
Matches 19; Conservative 17; Mismatches 22; Indels 4; Gaps 4;

Db 598 KYCRSK-YETIHGQNDNAADVLELAIKREMPAELL-R-ASLRHTNEDQNFLLNVGRSA 654
QY 43 KTCSCRLYELLHGAG-NHAAGILTLGKRSGPPGLQRLQRLQASGNHAAAGILTMGRRA 101

Db 655 SP 656
QY 102 GA 103

RESULT 7
ID 014514 PRELIMINARY; PRT; 1584 AA.
AC 014514;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE BAI 1.
GN BAI 1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=BRIN;
RA NISHIMORI H., SHIRATSUCHI T., URANO T., KIMURA Y., KIYONO K.,
RA TATSUMI K., YOSHIDA S., ONO M., KUWANO M., NAKAMURA Y.;
RL ONCOGENE 0:0-0(1997).
DR EMBL; AB005297; D1024528; -.
DR PFAM; PF00090; tsp_1; 5.
SQ SEQUENCE 1584 AA; 173531 MW; 235A5C42 CRC32;

Query Match 10.8%; Score 101; DB 4; Length 1584;
Best Local Similarity 50.0%; Pred. No. 2.81e-01;
Matches 16; Conservative 3; Mismatches 12; Indels 1; Gaps 1;

Db 12 WILAPLLLLLLGLRRARAAGADAGPGPEPC 43
QY 10 WAAVTLTLLLLLL-PPALLSSGAAQPLPDCC 40

RESULT 8
ID 060455 PRELIMINARY; PRT; 480 AA.
AC 060455;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLYADENYLATE BINDING PROTEIN-INTERACTING PROTEIN-1.
GN PAIP1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RC SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RX MEDLINE; 98208037.
RA CRAIG A.W.B., HAGHTGHAT A., YU A.T.K., SONENBERG N.;
RT "Interaction of polyadenylate-binding protein with the eIF4G

```

[illegible]



\*\*\*\*\*

W P E R E H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 20:52:28 1999; MasPar time 5.99 Seconds

Tabular output not generated. 618.331 Million cell updates/sec

Title: >US-08-938-548B-2

Description: (1-131) from US08938548B.pap

Perfect Score: 931

Sequence: 1 MNLPTKYSWAAVTLTLLLL.....GRRCSAPAAASVAPGGQSGI 131

Scoring table: PAM 150

Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot37

1:swissprot

Statistics: Mean 41.738; Variance 80.369; scale 0.519

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	109	11.7	602	1 PGH1_RAT	PROSTAGLANDIN G/H SYNT	2.91e-03
2	107	11.5	131	1 SECR_PIG	SECRETIN PRECURSOR (PR	5.66e-03
3	105	11.3	347	1 YFG-ECOLI	HYPOTHETICAL 38.7 KD P	1.90e-02
4	102	11.0	1744	1 TENS-CHICK	TENSIN.	2.90e-02
5	101	10.8	599	1 PGH1_HUMAN	PROSTAGLANDIN G/H SYNT	4.00e-02
6	99	10.6	1013	1 PTPX_MACNE	PROTEIN-TYROSINE PHOSP	7.56e-02
7	99	10.6	1015	1 PTPX_HUMAN	PROTEIN-TYROSINE PHOSP	7.56e-02
8	98	10.5	1165	1 CYA6_MOUSE	ADENYLATE CYCLASE, TYP	1.04e-01
9	97	10.4	260	1 URK1_MOUSE	URIDINE KINASE (EC 2.7	1.42e-01
10	97	10.4	602	1 PGH1_MOUSE	PROSTAGLANDIN G/H SYNT	1.42e-01
11	96	10.3	205	1 YK07_YEAST	HYPOTHETICAL 22.7 KD P	1.94e-01
12	94	10.1	101	1 GRO-GRIGR	GROWTH REGULATED PROTE	3.59e-01
13	94	10.1	292	1 Y152_HUMAN	HYPOTHETICAL PROTEIN K	3.59e-01
14	94	10.1	1337	1 PIP2_HUMAN	PROTEIN-TYROSINE PHOSP	3.59e-01
15	94	10.1	2499	1 MPRI_BOVIN	CATION-INDEPENDENT MAN	3.59e-01
16	93	10.0	1109	1 CYGD_CANFA	RETINAL GUANYLYL CYCIA	4.87e-01
17	92	9.9	235	1 FL3L_HUMAN	SL CYTOKINE PRECURSOR	6.60e-01
18	92	9.9	251	1 C1QB_HUMAN	COMPLEMENT C1Q SUBCOMP	6.60e-01
19	92	9.9	399	1 SHEG_PHOSU	SEX HORMONE-BINDING GL	6.60e-01
20	92	9.9	676	1 ICPO_HSVBK	TRANS-ACTING TRANSCRIP	6.60e-01
21	92	9.9	676	1 ICPO_HSVBK	TRANS-ACTING TRANSCRIP	6.60e-01
22	91	9.8	492	1 COGY_MOUSE	STROMELYSIN-3 PRECURSO	8.92e-01
23	89	9.6	76	1 CD24_MOUSE	SIGNAL TRANSDUCER CD24	1.62e+00

24	89	9.6	90	1 VGE_BPHX	LYSIS PROTEIN (E PROTE	1.62e+00
25	89	9.6	90	1 VGE_BPS13	LYSIS PROTEIN (E PROTE	1.62e+00
26	89	9.6	232	1 FL3L_MOUSE	SL CYTOKINE PRECURSOR	1.62e+00
27	89	9.6	322	1 RLUC_HAEN	RIBOSOMAL LARGE SUBUNI	1.62e+00
28	89	9.6	430	1 SNAK_DROME	SERINE PROTEASE SNAKE	1.62e+00
29	89	9.6	438	1 LCAT_MOUSE	PHOSPHATIDYLCHOLINE-ST	1.62e+00
30	88	9.5	238	1 EFA3_HUMAN	EPHRIN-A3 PRECURSOR (E	2.17e+00
31	88	9.5	317	1 LIP1_PSYIM	LIPASE 1 PRECURSOR (EC	2.17e+00
32	88	9.5	380	1 LEU3_PHACH	3-ISOPROPYLMALATE DEHY	2.17e+00
33	88	9.5	1165	1 CYA6_CANFA	ADENYLATE CYCLASE, TYP	2.17e+00
34	88	9.5	1166	1 CYA6_RAT	ADENYLATE CYCLASE, TYP	2.17e+00
35	87	9.3	229	1 PRL_FELCA	PROLACTIN PRECURSOR (P	2.90e+00
36	87	9.3	245	1 ICF3_HSV1N	INFECTED CELL PROTEIN	2.90e+00
37	87	9.3	319	1 RLUC_ECOLI	RIBOSOMAL LARGE SUBUNI	2.90e+00
38	87	9.3	419	1 HFLK_ECOLI	HFLK PROTEIN.	2.90e+00
39	87	9.3	600	1 PGH1_SHEEP	PROSTAGLANDIN G/H SYNT	2.90e+00
40	87	9.3	696	1 LSHR_PIG	LUTROPIN-CHORIOGNADOT	2.90e+00
41	87	9.3	2491	1 MPRI_HUMAN	CATION-INDEPENDENT MAN	2.90e+00
42	86	9.2	142	1 IL3_CALJA	INTERLEUKIN-3 PRECURSO	3.87e+00
43	86	9.2	370	1 WNT1_MOUSE	WNT-1 PROTO-ONCOGENE P	3.87e+00
44	86	9.2	1027	1 CAFF_RIFPA	FIBRIL-FORMING COLLAGE	3.87e+00
45	86	9.2	2194	1 SC16_YEAST	MULTIDOMAIN VESICLE CO	3.87e+00

## ALIGNMENTS

RESULT	1	PGH1_RAT	STANDARD;	PR	602 AA.
ID	PGH1_RAT	062731;	063684;		
AD	PGH1_RAT	062731;	063684;		
DT	15-DEC-1998	(REL. 37, CREATED)			
DT	15-DEC-1998	(REL. 37, LAST SEQUENCE UPDATE)			
DT	15-DEC-1998	(REL. 37, LAST ANNOTATION UPDATE)			
DE	PROSTAGLANDIN G/H SYNT	1 PRECURSOR (EC 1.14.99.1)	(CYCLOOXYGENASE		
DE	-1) (COX-1)	(PROSTAGLANDIN-ENDOROXIDE SYNTHASE 1)	(PROSTAGLANDIN H2		
DE	SYNTASE 1)	(PGH SYNTHASE 1) (PGHS-1) (PHS 1).			
GN	PTGS1 OR COX1 OR COX-1.				
OS	RATTUS NORVEGICUS (RAT).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;				
OC	RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SPRAGUE-DAWLEY;				
RX	MEDLINE; 94099619.				
RA	FENG L., SUN W., XIA Y., TANG W.W., CHANNUGAM P., SOYOOLA E.,				
RA	WILSON C.B., HWANG D.,				
RT	"Cloning two isoforms of rat cyclooxygenase: differential regulation				
RT	of their expression."				
RL	ARCH. BIOCHEM. BIOPHYS. 307:361-368(1993).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=FISHER 344; TISSUE=TRACHEA;				
RX	MEDLINE; 95168876.				
RA	KITZLER J., HILL E., HARDMAN R., REDDY N., PHILPOT R., ELING T.E.;				
RT	"Analysis and quantitation of splicing variants of the TPA-inducible				
RT	PGHS-1 mRNA in rat tracheal epithelial cells."				
RL	ARCH. BIOCHEM. BIOPHYS. 316:856-863(1995).				
CC	-!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN REGULATING OR PROMOTING				
CC	CELL PROLIFERATION IN SOME NORMAL AND NEOPLASTICALLY TRANSFORMED				
CC	CELLS.				
CC	-!- CATALYTIC ACTIVITY: ARACHIDONATE + AH(2) + 2 O(2) = PROSTAGLANDIN				
CC	H2 + A + H(2)O.				
CC	-!- PATHWAY: FIRST STEP IN THE FORMATION OF PROSTAGLANDINS AND				
CC	THROMBOXANES.				
CC	-!- SUBUNIT: HOMODIMER (BY SIMILARITY).				
CC	-!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MICROSOMAL MEMBRANE.				
CC	-!- THIS ENZYME ACTS BOTH AS A DIOXYGENASE AND AS A PEROXIDASE.				
CC	-!- THIS ENZYME IS THE TARGET OF NONSTEROIDAL ANTI-INFLAMMATORY DRUGS				
CC	SUCH AS ASPIRIN.				
CC	-!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.				
CC	-!- SIMILARITY: BELONGS TO THE PROSTAGLANDIN G/H SYNTHASE FAMILY.				

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

DR PFAM; PF00123; hormone2; 1.
DR HSP; P01274; 1GDN.
KW GLUCAGON FAMILY; HORMONE; AMIDATION; SIGNAL.
FT NON_TER 1
FT SIGNAL <1 18
FT PEPTIDE 30 56
FT MOD_RES 56
SQ SEQUENCE 131 AA; 14277 MW; 837D201A CRC32;

Query Match 11.5%; Score 107; DB 1; Length 131;
Best Local Similarity 72.7%; Pred. No. 5.66e-03;
Matches 16; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 3 LLLLLLPPPLLLAGCAARPAP 24
   | | | | | | | | | | | | | | | |
QY 16 LLLLLLPPPLSSGAAQPLP 37

RESULT 3
ID YPFG_ECOLI STANDARD; PRT; 347 AA.
AC F76539;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 38.7 KD PROTEIN IN TKIB-NARQ INTERGENIC REGION PRECURSOR.
GN YPFG.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL SCIENCE 277:1453-1474(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000333; G1788809; -
CC ECOGENE; EG14194; YPFG.
KW HYPOTHETICAL PROTEIN; SIGNAL.
FT SIGNAL 1 21
FT CHAIN 22 347
SQ SEQUENCE 347 AA; 38746 MW; 136C8304 CRC32;

Query Match 11.3%; Score 105; DB 1; Length 347;
Best Local Similarity 25.0%; Pred. No. 1.09e-02;
Matches 17; Conservative 29; Mismatches 20; Indels 2; Gaps 1;

Db 7 LFFFFALLPTSLVWAAPARAFSD - WQVTNNQNFCAVNTGDHNGLVMTLSRSAGHT 64
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 15 LLLLLLLPALSSGAAQPLDCCQKTCRSLRYELLHGAGNHAAGILTIKRRSGPP 74

Db 65 DAVIRIER 72
   | | | |
QY 75 GLOGRLQR 82

RESULT 4
ID TENS_CHICK STANDARD; PRT; 1744 AA.
AC Q04205; Q91007; Q92011;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

```

---

```

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE TENSIN.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RX MEDLINE; 94350987.
RA LO S.H., AN Q., BAO S., WONG W.K., LIU Y., JANNEY P.A., HARTWIG J.H.,
RA CHEN L.B.;
RT "Molecular cloning of chick cardiac muscle tensin. Full-length cDNA
sequence, expression, and characterization.";
RL J. BIOL. CHEM. 269:22310-22319(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RX MEDLINE; 95204530.
RA CHUANG J.Z., LIN D.C., LIN S.;
RT "Molecular cloning, expression, and mapping of the high affinity
actin-capping domain of chicken cardiac tensin.";
RL J. CELL BIOL. 128:1095-1109(1995).
RN [3]
RP SEQUENCE FROM N.A.
RA CHEN L.B.;
RL SUBMITTED (XXX-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE OF 1469-1744 FROM N.A.
RC TISSUE=EMBRYONIC CHONDROCYTES, AND EMBRYONIC HEART;
RX MEDLINE; 94039118.
RA VAN DE WERKEN R., GENNARI M., TAVELLA S., BET P., MOLINA F.,
RA LIN S., CANCEDDA R., CASTAGNOLA P.;
RT "Modulation of tensin and vimentin expression in chick embryo
developing cartilage and cultured differentiating chondrocytes.";
RL EUR. J. BIOCHEM. 217:781-790(1993).
RN [5]
RP SH2 DOMAIN.
RX MEDLINE; 91220073.
RA DAVIS S., LU M.L., LO S.H., LIN S., BUTLER J.A., DRUKER B.J.,
RA ROBERTS T.M., AN Q., CHEN L.B.;
RT "Presence of an SH2 domain in the actin-binding protein tensin.";
RL SCIENCE 252:712-715(1991).
CC -1- FUNCTION: MAY BE INVOLVED IN CELL MIGRATION, CARTILAGE DEVELOPMENT
CC AND IN LINKING SIGNAL TRANSDUCTION PATHWAYS TO THE CYTOSKELETON.
CC BINDS TO ACTINS AND PHOSPHORYLATED PROTEINS IN SRC-TRANSFORMED
CC CELLS. MAY BIND ACTIN WITH CAPPING AND BUNDLING PROPERTIES.
CC -1- SUBCELLULAR LOCATION: LOCALIZES TO ADHERENS JUNCTIONS.
CC -1- TISSUE SPECIFICITY: HEART, GIZZARD, LUNG AND SKELETAL MUSCLE.
CC -1- PTM: TYROSINE-PHOSPHORYLATED.
CC -1- SIMILARITY: CONTAINS A TENSIN DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M96625; G619577; -
CC EMBL; L06662; G212755; ALT_INIT.
CC EMBL; Z18529; G63805; ALT_INIT.
CC EMBL; M74165; G212752; -
CC EMBL; X66286; G63803; -
CC PROSITE; PS50001; SH2; 1.
CC PFAM; PF00017; SH2; 1.
CC HSP; P16277; IBLJ.
KW ACTIN-BINDING; CYTOSKELETON; SH2 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 66 342
FT TENSIN.
FT SH2.
FT CONFLICT 49 49 C -> R (IN REF. 2).
FT CONFLICT 61 61 M -> T (IN REF. 2).

```

SQ SEQUENCE 599 AA; 68656 MW; 8C7684CD CRC32;  
 Query Match 10.8%; Score 101; DB 1; Length 599;  
 Best Local Similarity 51.9%; Pred. No. 4.00e-02;  
 Matches 14; Conservative 5; Mismatches 7; Indels 1; Gaps 1;

Db 10 LFLFLLPPLVLLADPGATPVNCC 36  
 ||:|||||:|:|:|:|:|:|:|:|:|  
 QY 15 LLLLLLLP-FALLSSGAAQPLDCC 40

RESULT 6  
 ID PTPX\_MACNE STANDARD; PRT; 1013 AA.  
 AC O02695;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE PROTEIN-TYROSINE PHOSPHATASE X PRECURSOR (EC 3.1.3.48) (R-PTP-X)  
 DE (M1851).  
 GN PTPRN2.  
 OS MACACA NEMESTRINA (PIG-TAILED MACAQUE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; MACACA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PANCREATIC ISLETS;  
 RX MEDLINE; 97254813.  
 RA LAGASSE J., JELINEK L., SEXSON S., LOFTON-DAY C., BREININGER J.,  
 RA SHEPPARD P., KINDSVOGEL W., HAGOPIAN W.A.;  
 RT "An islet-cell protein tyrosine phosphatase is a likely precursor to  
 RT the 37-kDa autoantigen in type 1 diabetes: human and macaque  
 RT sequences, tissue distribution, unique and shared epitopes, and  
 RT predictive autoantibodies.";  
 RL MOL. MED. 3:163-173(1997).  
 CC -!- FUNCTION: IMPLICATED IN DEVELOPMENT OF NERVOUS SYSTEM AND  
 CC PANCREATIC ENDOCRINE CELLS.  
 CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATASE + H(2)O =  
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE).  
 CC -!- TISSUE SPECIFICITY: BRAIN, PROSTATE, PANCREATIC ISLETS. LOWER  
 CC EXPRESSION IN SPINAL CORD, THYROID, ADRENAL MEDULLA AND  
 CC GASTROINTESTINAL TRACT.  
 CC -!- PTM: APPEARS TO UNDERGO MULTIPLE PROTEOLYTIC CLEAVAGE AT  
 CC CONSECUTIVE BASIC RESIDUES (BY SIMILARITY).  
 CC -!- SIMILARITY: CONTAINS 1 PROTEIN-TYROSINE PHOSPHATASE DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U91574; GI916942; -.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
 DR PFAM; PF00102; Y\_phosphatase; 1.  
 DR HSSP; P18052; LYFO.  
 KW HYDROLASE; RECEPTOR; GLYCOPROTEIN; SIGNAL; TRANSMEMBRANE.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 1013 PROTEIN-TYROSINE PHOSPHATASE X.  
 FT DOMAIN 20 613 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 614 634 POTENTIAL.  
 FT DOMAIN 635 1013 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 768 1002 PROTEIN-TYROSINE PHOSPHATASE.  
 FT ACT\_SITE 943 943 POTENTIAL.  
 FT SITE 425 426 CLEAVAGE SITE (BY SIMILARITY).  
 FT CARBOHYD 562 562 POTENTIAL.  
 SQ SEQUENCE 1013 AA; 111190 MW; 9EC7D263 CRC32;

Query Match 10.6%; Score 99; DB 1; Length 1013;

Best Local Similarity 57.1%; Pred. No. 7.56e-02;  
 Matches 12; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 6 LLLLLLLPRLVLPAPSSVP 26  
 |||||:|:|:|:|:|:|:|:|:|  
 QY 15 LLLLLLLP-FALLSSGAAQAP 35

RESULT 7  
 ID PTPX\_HUMAN STANDARD; PRT; 1015 AA.  
 AC Q92932; Q92662;  
 DT 01-NOV-1997 (REL. 35, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE PROTEIN-TYROSINE PHOSPHATASE X PRECURSOR (EC 3.1.3.48) (R-PTP-X)  
 DE (ISLET CELL AUTOANTIGEN RELATED PROTEIN) (ICAR) (PHOGRIN).  
 GN PTPRN2.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PANCREAS;  
 RX MEDLINE; 97032784.  
 RA KAWASAKI E., HUTTON J.C., EISENBARTH G.S.;  
 RA "Molecular cloning and characterization of the human transmembrane  
 RA protein tyrosine phosphatase homologue, phogrin, an autoantigen of  
 RA type 1 diabetes.";  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 227:440-447(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-FETAL BRAIN;  
 RX MEDLINE; 97117415.  
 RA SMITH P.D., BARKER K.T., WANG J., LU Y.-J., SHIPLEY J., CROMPTON M.R.;  
 RA "ICAR, a novel member of a new family of transmembrane, tyrosine  
 RA phosphatase-like proteins.";  
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 229:402-411(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN, AND PANCREAS;  
 RX MEDLINE; 96394649.  
 RA CUI L., YU W.-P., DE AIZPURUA H.J., SCHMIDT R.S., PALLEN C.J.;  
 RA "Cloning and characterization of islet cell antigen-related protein-  
 RA tyrosine phosphatase (PTP), a novel receptor-like PTP and autoantigen  
 RA in insulin-dependent diabetes.";  
 RL J. BIOL. CHEM. 271:24817-24823(1996).  
 CC -!- FUNCTION: IMPLICATED IN DEVELOPMENT OF NERVOUS SYSTEM AND  
 CC PANCREATIC ENDOCRINE CELLS. OPTIMUM ACTIVITY IS MEASURED AT PH  
 CC 4.5.  
 CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =  
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE).  
 CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS IN BRAIN AND PANCREAS. LOWER  
 CC LEVELS IN TRACHEA, PROSTATE, STOMACH AND SPINAL CHORD.  
 CC -!- PTM: APPEARS TO UNDERGO MULTIPLE PROTEOLYTIC CLEAVAGE AT  
 CC CONSECUTIVE BASIC RESIDUES.  
 CC -!- DOMAIN: THE CYTOPLASMIC DOMAIN APPEARS TO CONTAIN THE  
 CC AUTOANTIGENIC EPITOPES.  
 CC -!- DISEASE: AUTOANTIGEN IN INSULIN-DEPENDENT DIABETES MELLITUS  
 CC (IDDM).  
 CC -!- SIMILARITY: CONTAINS 1 PROTEIN-TYROSINE PHOSPHATASE DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U66702; G1620664; -.  
 DR EMBL; Y08569; E273864; -.  
 DR EMBL; AF007555; G2262075; -.

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: L31783; G471981; -  
 DR MGD; MGI:98904; UNPK.  
 DR TRANSFERASE; KINASE.  
 FT NON\_TER 1  
 SQ SEQUENCE 260 AA; 29622 MW; E72BB622 CRC32;

Query Match 10.4%; Score 97; DB 1; Length 260;  
 Best Local Similarity 34.1%; Pred. No. 1.42e-01;  
 Matches 15; Conservative 14; Mismatches 13; Indels 2; Gaps 2;

Db 218 LCKRRGGNGNRH-KTEPEPGDHP-GVLATGKRSHLESSRP 259

QY 66 LKRRSGPGGLQRLQLQASGHAAGILTMGRRAEPAPRP 109

RESULT 10  
 ID PGH1\_MOUSE STANDARD; PRT; 602 AA.  
 AC P22437; (REL. 19, CREATED)  
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE PROSTAGLANDIN G/H SYNTHASE 1 PRECURSOR (EC 1.14.99.1) (CYCLOOXYGENASE  
 DE -1) (COX-1) (PROSTAGLANDIN-ENDOPEROXIDE SYNTHASE 1) (PROSTAGLANDIN H2  
 DE SYNTHASE 1) (PGH SYNTHASE 1) (PGHS-1) (PHS 1).  
 GN PTGS1 OR COX1 OR COX-1.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 90203007.  
 RA DEWITT D.L., EL-HARITH E.A., KRAEMER S.A., ANDREWS M.J., YAO E.F.,  
 RA ARMSTRONG R.L., SMITH W.L.;  
 RT "The aspirin and heme-binding sites of ovine and murine prostaglandin  
 RT endoperoxide synthases."  
 RL J. BIOL. CHEM. 265:5192-5198(1990).  
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN REGULATING OR PROMOTING  
 CC CELL PROLIFERATION IN SOME NORMAL AND NEOPLASTICALLY TRANSFORMED  
 CC CELLS.  
 CC -1- CATALYTIC ACTIVITY: ARACHIDONATE + AH(2) + 2 O(2) = PROSTAGLANDIN  
 CC H2 + A + H(2)O.  
 CC -1- PATHWAY: FIRST STEP IN THE FORMATION OF PROSTAGLANDINS AND  
 CC THROMBOXANES.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MICROSOMAL MEMBRANE.  
 CC -1- THIS ENZYME ACTS BOTH AS A DIOXYGENASE AND AS A PEROXIDASE.  
 CC -1- THIS ENZYME IS THE TARGET OF NONSTEROIDAL ANTI-INFLAMMATORY DRUGS  
 CC SUCH AS ASPIRIN.  
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE PROSTAGLANDIN G/H SYNTHASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M34141; G200303; -  
 DR MGD; MGI:97797; PTGS1.  
 DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE; PS01186; EGF\_2; FALSE\_NEG.  
 DR PFAM; PF00008; EGF; 1.  
 DR HSP; P05979; 1PRH.  
 KW OXIDOREDUCTASE; DIOXYGENASE; PEROXIDASE; GLYCOPROTEIN; ACETYLATION;  
 KW PROSTAGLANDIN BIOSYNTHESIS; HEME; IRON; SIGNAL; MEMBRANE;  
 KW EGF-LIKE DOMAIN.  
 FT SIGNAL 1 26

FT CHAIN 27 602  
 FT DOMAIN 34 72  
 FT ACT\_SITE 209 387  
 FT ACT\_SITE 387 387  
 FT BINDING 390 390  
 FT MOD\_RES 532 532  
 FT DISULFID 38 49  
 FT DISULFID 43 59  
 FT DISULFID 61 71  
 FT DISULFID 39 161  
 FT DISULFID 571 577  
 FT CARBOHYD 70 70  
 FT CARBOHYD 106 106  
 FT CARBOHYD 146 146  
 SQ SEQUENCE 602 AA; 69042 MW; 96489281 CRC32;

Query Match 10.4%; Score 97; DB 1; Length 602;  
 Best Local Similarity 38.7%; Pred. No. 1.42e-01;  
 Matches 12; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

Db 9 WFLPLLLLLPTPSVLLADPGVSPVNPCC 39

QY 10 WAAVTLLELLLLPALLSGAAQPLPDC 40

RESULT 11  
 ID YK07\_YEAST STANDARD; PRT; 205 AA.  
 AC P36061;  
 DT 01-JUN-1994 (REL. 29, CREATED)  
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 22.7 KD PROTEIN IN SDH5/YTA3 INTERGENIC REGION.  
 GN YKL147C OR YKL601.  
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;  
 OC SACCHAROMYCETACEAE; SACCHAROMYCES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C;  
 RX MEDLINE; 94378720.  
 RA VANDENBOL M., BOLLE P.-A., DION C., PORTELETTE D., HILGER F.;  
 RT "DNA sequencing of a 36.2 kb fragment located between the FAS1 and  
 RT LAP loci of chromosome XI of Saccharomyces cerevisiae."  
 RL YEAST 10:S35-S40(1994).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Z26877; G407501; -  
 DR EMBL; Z28146; G486251; -  
 DR PIR; S37804; S37804.  
 DR PIR; S44581; S44581.  
 KW HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 205 AA; 22673 MW; 0AEABD4E CRC32;

Query Match 10.3%; Score 96; DB 1; Length 205;  
 Best Local Similarity 53.8%; Pred. No. 1.94e-01;  
 Matches 14; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Db 74 LLLLLLLLLLLLLPLPSVKGEPPAC 99

QY 15 LLLLLLLLLLLLLSGAAQPLPDC 40

RESULT 12  
 ID GRO\_CRIGR STANDARD; PRT; 101 AA.  
 AC P09340;  
 DT 01-MAR-1989 (REL. 10, CREATED)

FT	DOMAIN	366	446	FIBRONECTIN TYPE-III.
FT	DOMAIN	454	532	FIBRONECTIN TYPE-III.
FT	DOMAIN	540	615	FIBRONECTIN TYPE-III.
FT	DOMAIN	626	710	FIBRONECTIN TYPE-III.
FT	DOMAIN	1065	1337	PROTEIN-TYROSINE PHOSPHATASE.
FT	ACT_SITE	1239	1239	BY SIMILARITY.
FT	CARBOHYD	72	72	POTENTIAL.
FT	CARBOHYD	82	82	POTENTIAL.
FT	CARBOHYD	93	93	POTENTIAL.
FT	CARBOHYD	104	104	POTENTIAL.
FT	CARBOHYD	142	142	POTENTIAL.
FT	CARBOHYD	172	172	POTENTIAL.
FT	CARBOHYD	192	192	POTENTIAL.
FT	CARBOHYD	231	231	POTENTIAL.
FT	CARBOHYD	258	258	POTENTIAL.
FT	CARBOHYD	278	278	POTENTIAL.
FT	CARBOHYD	342	342	POTENTIAL.
FT	CARBOHYD	351	351	POTENTIAL.
FT	CARBOHYD	376	376	POTENTIAL.
FT	CARBOHYD	391	391	POTENTIAL.
FT	CARBOHYD	396	396	POTENTIAL.
FT	CARBOHYD	413	413	POTENTIAL.
FT	CARBOHYD	431	431	POTENTIAL.
FT	CARBOHYD	501	501	POTENTIAL.
FT	CARBOHYD	525	525	POTENTIAL.
FT	CARBOHYD	536	536	POTENTIAL.
FT	CARBOHYD	582	582	POTENTIAL.
FT	CARBOHYD	603	603	POTENTIAL.
FT	CARBOHYD	618	618	POTENTIAL.
FT	CARBOHYD	628	628	POTENTIAL.
FT	CARBOHYD	637	637	POTENTIAL.
FT	CARBOHYD	666	666	POTENTIAL.
FT	CARBOHYD	669	669	POTENTIAL.
FT	CARBOHYD	761	761	POTENTIAL.
FT	CARBOHYD	772	772	POTENTIAL.
FT	CARBOHYD	784	784	POTENTIAL.
FT	CARBOHYD	790	790	POTENTIAL.
FT	CARBOHYD	824	824	POTENTIAL.
FT	CARBOHYD	910	910	POTENTIAL.
FT	CARBOHYD	937	937	POTENTIAL.
FT	SEQUENCE	1337 AA;	145985 MW; 55F90A6B CRC32;	
DB	Query Match	10.18;	Score 94;	DB 1; Length 1337;
QY	Best Local Similarity	48.38;	Pred. No. 3.59e-01;	
	Matches 14:	Conservative	7; Mismatches	6; Indels 2; Gaps 2;
		19 WA-LPLLLLLLRLGQ-ILCAGTGTPSPDP 45		
		:       :    :    :		
		10 WAAVTLTLLLLPPALLSSGAAQPLPD 38		
RESULT	15			
ID	MPRI_BOVIN	STANDARD;	PRT;	2499 AA.
AC	F08169;			
DT	01-AUG-1988 (REL. 08, CREATED)			
DT	01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)			
DT	01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)			
DE	CATION-INDEPENDENT MANNOSE-6-PHOSPHATE RECEPTOR PRECURSOR (CI MAN-6-P			
DE	RECEPTOR) (CI-MPR) (INSULIN-LIKE GROWTH FACTOR II RECEPTOR).			
GN	IGF2R.			
OS	BOS TAURUS (BOVINE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;			
OC	ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 88115411.			
RA	LOBEL P., DAHMS N.M., KORNFIELD S.;			
RT	"Cloning and sequence analysis of the cation-independent mannose 6-			
RT	phosphate receptor."			
RL	J. BIOL. CHEM. 263:2563-2570(1988).			
RN	[2]			
RN	SEQUENCE OF 1039-2499 FROM N.A.			
RP	MEDLINE; 87175648.			

\*\*\*\*\*

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Run on: Fri Aug 20 20:51:49 1999; MasPar time 8.66 Seconds

Tabular output not generated. 506.456 Million cell updates/sec

Title: >US-08-938-548B-2  
Description: (1-131) from US08938548B.pap  
Perfect Score: 931  
Sequence: 1 MNLPTKVSAAVTLTLLLL.....GRRCSAPAAASVAPGGQSGI 131

Scoring table: PAM 150  
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 40.416; Variance 87.828; scale 0.460

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	110	11.8	1321	2	T00382	hypothetical protein
2	109	11.7	602	2	S69198	prostaglandin G/H syn
3	108	11.6	599	2	A36746	prostaglandin-endoper
4	107	11.5	131	1	SEPG	secretin precursor -
5	107	11.5	602	2	S39782	cyclooxigenase 1 - ra
6	105	11.3	347	2	A65022	hypothetical protein
7	103	11.1	679	2	C71413	hypothetical protein
8	103	11.1	1792	2	A57075	tensin - chicken (fra
9	102	11.0	1733	2	S27939	tensin - chicken
10	102	11.0	1744	2	A54970	tensin, cardiac muscl
11	101	10.8	562	2	A38146	prostaglandin-endoper
12	101	10.8	599	2	JH0259	prostaglandin-endoper
13	101	10.8	1584	2	T00026	brain-specific anglog
14	99	10.6	245	2	S43293	phogrin precursor - h
15	99	10.6	1015	2	JC5062	transmembrane tyrosin
16	99	10.6	1015	2	JC5263	adenylate cyclase (EC
17	98	10.5	1166	2	A49201	prostaglandin-endoper
18	97	10.4	602	2	A35564	hypothetical protein
19	96	10.3	205	2	S37804	hypothetical protein
20	96	10.3	312	2	A61183	hypothetical protein
21	95	10.2	226	2	A70565	probable cutinase pre
22	95	10.2	491	2	JC6197	stromelysin 3 (EC 3.4
23	94	10.1	101	2	B28414	growth-regulated prot

24 94 10.1 1337 1 I38670 protein-tyrosine-phos 1.03e+00  
25 94 10.1 1736 2 T00391 hypothetical protein 1.03e+00  
26 94 10.1 2499 2 A30788 mannose 6-phosphate r 1.03e+00  
27 92 9.9 178 2 I39076 FLT3 ligand alternati 1.79e+00  
28 92 9.9 235 2 I38440 FLT3 ligand - human 1.79e+00  
29 92 9.9 235 2 S43292 FLT3/FLK2 ligand (clo 1.79e+00  
30 92 9.9 253 1 C1HU0B complement subcompone 1.79e+00  
31 92 9.9 676 1 EDBE22 immediate-early prote 1.79e+00  
32 92 9.9 676 1 EDBE23 immediate-early prote 1.79e+00  
33 91 9.8 492 2 A44399 stromelysin 3 (EC 3.4 2.36e+00  
34 90 9.7 147 2 S24303 hypothetical protein 3.10e+00  
35 89 9.6 76 2 A43537 heat-stable antigen M 4.06e+00  
36 89 9.6 90 1 ZEBPF4 gene E protein - phag 4.06e+00  
37 89 9.6 91 2 S47060 lysis protein - phage 4.06e+00  
38 89 9.6 91 2 JS0455 gene E protein - phag 4.06e+00  
39 89 9.6 220 2 S43291 FLT3/FLK2 ligand (clo 4.06e+00  
40 89 9.6 220 2 I58343 FLT3 ligand isoform 5 4.06e+00  
41 89 9.6 231 2 A49265 FLT3/flk-2 ligand pre 4.06e+00  
42 89 9.6 322 2 G84151 hypothetical protein 4.06e+00  
43 89 9.6 430 2 A24702 serine proteinase sna 4.06e+00  
44 89 9.6 438 1 XXMSN phosphatidylcholine-- 4.06e+00  
45 88 9.5 317 2 S28225 triacylglycerol lipas 5.32e+00

## ALIGNMENTS

RESULT 1  
ENTRY T00382 #type fragment  
TITLE hypothetical protein KIAA0634 - human (fragment)  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 01-Feb-1999  
ACCESSIONS T00382  
REFERENCE Z14142  
#authors Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.  
#journal DNA Res. (1998) 5:169-176  
#title Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.  
#accession T00382  
#status preliminary; translated from GB/EMBL/DBJ  
#molecule\_type mRNA  
#residues 1-1321 #label ISH  
#cross-references EMBL:AB014534; NID:d1204289; PID:d1032570  
#experimental\_source brain  
GENETICS KIAA0634  
#note #length 1321 #checksum 8342  
SUMMARY  
Query Match 11.8%; Score 110; DB 2; Length 1321;  
Best Local Similarity 57.6%; Pred. No. 9.52e-03;  
Matches 19; Conservative 7; Mismatches 4; Indels 3; Gaps 3;  
Db 68 LLLFLLPPPPPLAGATAAASREPDSPCLKLT 100  
QY 15 LLLLLLLLPPA-LLSSGAAQPL-PDC-CRQKT 44

RESULT 2  
ENTRY S69198 #type complete  
TITLE prostaglandin G/H synthase 1 - rat  
ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat  
DATE 24-Aug-1996 #sequence\_revision 13-Mar-1997 #text\_change 10-Sep-1997  
ACCESSIONS S69198; S69199  
REFERENCE S69198  
#authors Kitzler, J.W.  
#submission submitted to the EMBL Data Library, December 1994  
#accession S69198  
#status preliminary  
#molecule\_type mRNA  
#residues 1-602 #label KIT

```

Matches 15; Conservative 6; Mismatches 5; Indels 2; Gaps 1;

Db 12 LLLLLLLLLPPPPVLLLDAGVSPVIPC 39
| | | | | | | | | | | | | | | | | |
QY 15 LLLLLLLLLPPA--LLSGAAQPLDCC 40

RESULT 6
ENTRY A65022 #type complete
TITLE hypothetical protein b2466 - Escherichia coli (strain K-12)
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
14-Nov-1997

ACCESSIONS A65022
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.

#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession A65022
#status preliminary; nucleic acid sequence not shown;
translation not shown

#molecule_type DNA
#residues 1-347 #label BLAT
#cross-references GB:A6000333; GB:U00096; NID:q1788805; PID:q1788809;
#experimental_source strain K-12, substrain MG1655
SUMMARY #length 347 #molecular_weight 38746 #checksum 7897

Query Match 11.3%; Score 105; DB 2; Length 347;
Best Local Similarity 25.08; Pred. No. 4.29e-02;
Matches 17; Conservative 29; Mismatches 20; Indels 2; Gaps 1;

Db 7 LLEFFALLPSLYWAAPARAFSD--WQVTCNNQNFVARTGHNGLVMTLSRGAHT 64
| | | | | | | | | | | | | | | | | |
QY 15 LLLLLLLPALLSGAAQPLDCCQKTCSCRLYELLHGAGNHAAGILTLGRRSQPP 74

Db 65 DAVURIER 72
| | | |
QY 75 GLOGRLQR 82

RESULT 7
ENTRY C71413 #type complete
TITLE hypothetical protein - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
#variety columbia
DATE 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change
05-Dec-1998
C71413
ACCESSIONS A71400
REFERENCE A71400
#authors Beran, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.;
Dean, C.; Bergkamp, R.; Dirksey, W.; Van Staveren, M.;
Stiekema, W.; Drost, L.; Ridley, P.; Hudson, S.A.; Patel,
K.; Murphy, G.; Piffanelli, P.; Wedler, H.; Wedler, E.;
Wambutt, R.; Weitzengraber, T.; Pohl, T.M.; Terry, N.;
Gielen, J.; Villarroel, R.; De Clerck, R.; Van Montagu, M.;
Leclercq, A.; Auborg, S.; Gy, I.; Kreis, M.; Lao, N.;
Kavanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger,
M.; Schaeffer, M.; Funk, B.; Mueller-Auer, S.; Silvey, M.;
James, R.; Montfort, A.; Pons, A.; Puigdomenech, P.; Douka,
A.; Vouklatou, E.; Milloni, D.; Hatzopoulos, P.;
Piravandi, E.; Obermaier, B.; Hilbert, H.; Duesterhoft, A.;
Moore, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.;
Rechman, S.; Ansoorge, W.; Cooke, R.; Berger, C.; Delseny,
M.; Voet, M.; Volckaert, G.; Mewes, H.W.; Klosterman, S.;
Schueller, C.; Chludzisz, N.
#journal Nature (1998) 391:485-488

```

```

#title Analysis of 1.9 Mb of contiguous sequence from chromosome 4
of Arabidopsis thaliana.
#cross-references MUID:98121113
#accession C71413
#status preliminary; nucleic acid sequence not shown;
translation not shown

#molecule_type DNA
#residues 1-679 #label BEV
#cross-references GB:297337; NID:g2244829; PID:e326841; PID:g2244855
GENETICS
#map_position 4COP9-4G3845
SUMMARY #length 679 #molecular_weight 74635 #checksum 9028

Query Match 11.1%; Score 103; DB 2; Length 679;
Best Local Similarity 30.6%; Pred. No. 7.76e-02;
Matches 19; Conservative 17; Mismatches 22; Indels 4; Gaps 4;

Db 598 KYCRSK-YETIHQNDNAADVLEIAIKREMPAELL-R-ASLRHTNEDORNFLNVGRSA 654
| | | | | | | | | | | | | | | | | |
QY 43 KTCRLYELLHGAG-NHAAGILTLGRRSQPPGLQRLQLQASGNHAGILTMGRRRA 101

Db 655 SP 656
| |
QY 102 GA 103

RESULT 8
ENTRY A57075 #type fragment
TITLE tensin - chicken (fragment)
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change
12-Feb-1999
ACCESSIONS A57075
REFERENCE A57075
#authors Chuang, J.Z.; Lin, D.C.; Lin, S.
#journal J. Cell Biol. (1995) 128:1095-1109
#title Molecular cloning, expression, and mapping of the high
affinity actin-capping domain of chicken cardiac tensin.
#cross-references MUID:95204530
#accession A57075
#status preliminary
#molecule_type mRNA
#residues 1-1792 #label CHU
#cross-references GB:I06662; NID:g212754; PID:g212755
CLASSIFICATION #superfamily SH2 homology
FEATURE 1520-1629 #domain SH2 homology #label SH2
SUMMARY #length 1792 #checksum 2643

Query Match 11.1%; Score 103; DB 2; Length 1792;
Best Local Similarity 38.1%; Pred. No. 7.76e-02;
Matches 24; Conservative 13; Mismatches 22; Indels 4; Gaps 4;

Db 1287 RAGFQDPQARQQPOVSVGVGHALPSPRTLHRTVATNTPSPGFGRRANPAVASV-PG 1345
| | | | | | | | | | | | | | | | | |
QY 70 RSG-PPGLQRLQLQASGNHA-AGIL-TMGRRAGAEAPRCLGRRCSSAPAAASVAPG 126

Db 1346 SPG 1348
| |
QY 127 GQS 129

RESULT 9
ENTRY S27939 #type complete
TITLE tensin - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
12-Feb-1999
ACCESSIONS S27939; S28973
REFERENCE S27939
#authors Chen, L.B.
#submission submitted to the EMBL Data Library, August 1991
#accession S27939

```

```

REFERENCE A39937
#authors Funk, C.D.; Funk, L.B.; Kennedy, M.E.; Pong, A.S.;
#journal Fitzgerald, G.A.
#title FASEB J. (1991) 5:2304-2312
#cross-references Human platelet/erythroleukemia cell prostaglandin G/H
#accession assignment. cDNA cloning, expression, and gene chromosomal
#molecule_type mRNA
#residues 1-1584 #label NIS
#cross-references EMBL:AB005297; NID:d1175078; PID:d1024528
#experimental_source brain
GENETICS
#gene GDB:BA11
#cross-references GDB:9838088; OMIM:602682
#map_position 8q24-8q24
CLASSIFICATION #superfamily thrombospondin type 1 repeat homology
FEATURE 408-462 #domain thrombospondin type 1 repeat homology #label
SUMMARY #length 1584 #molecular-weight 173531 #checksum 7909
Query Match 10.8%; Score 101; DB 2; Length 1584;
Best Local Similarity 50.0%; Pred. No. 1.39e-01;
Matches 16; Conservative 3; Mismatches 12; Indels 1; Gaps 1;
Db 12 WILAPLLLLLLGRRARAAAGADAGPGPEPC 43
QY 10 WAAVTLTLLLLLL-PPALLSSGAAQPLPDC 40

RESULT 14
ENTRY S43293 #type complete
TITLE FLT3/FLK2 ligand (clone S109) - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 20-Oct-1994 #sequence_revision 10-Nov-1995 #text_change
17-Mar-1999
ACCESSIONS S43293
REFERENCE S43290
#authors Hannun, C.; Culpepper, J.; Campbell, D.; McClanahan, T.;
Zurawski, S.; Bazan, J.F.; Kastelein, R.; Hudak, S.;
Wagner, J.; Mattson, J.; Luh, J.; Duda, G.; Martina, N.;
Peterson, D.; Menon, S.; Shanafelt, A.; Muench, M.; Kellner,
G.; Namikawa, R.; Rennick, D.; Roncarolo, M.G.; Zlotnik,
A.; Rosnet, O.; Dubreuil, P.; Birnbaum, D.; Lee, F.
#journal Nature (1994) 368:643-648
#title Ligand for FLT3/FLK2 receptor tyrosine kinase regulates
growth of haematopoietic stem cells and is encoded by
variant RNAs.
#cross-references MUID:94195428
#accession S43293
#status preliminary
#molecule_type mRNA
#residues 1-245 #label HAN
#note the authors translated the codon AGT for residue 25 as
Met
SUMMARY #length 245 #molecular-weight 27404 #checksum 295
Query Match 10.6%; Score 99; DB 2; Length 245;
Best Local Similarity 47.6%; Pred. No. 2.49e-01;
Matches 10; Conservative 9; Mismatches 2; Indels 0; Gaps 0;
Db 7 AWSPTYLLLLLLSSGLMGT 27
QY 9 SWAAVTLTLLLLPPALLSS 29

RESULT 15
ENTRY JCS062 #type complete
TITLE phogrin precursor - human
CONTAINS protein-tyrosine-phosphatase (EC 3.1.3.48) receptor type
ORGANISM #formal_name Homo sapiens #common_name man
DATE 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change
13-Nov-1998
ACCESSIONS JCS062
REFERENCE JCS062
#authors Kawasaki, E.; Hutton, J.C.; Eisenbarth, G.S.

```

---

```

REFERENCE A39937
#authors Funk, C.D.; Funk, L.B.; Kennedy, M.E.; Pong, A.S.;
#journal Fitzgerald, G.A.
#title FASEB J. (1991) 5:2304-2312
#cross-references Human platelet/erythroleukemia cell prostaglandin G/H
#accession assignment. cDNA cloning, expression, and gene chromosomal
#molecule_type mRNA
#residues 1-599 #label FUN
#cross-references GB:M59979; NID:g189886; PID:g189887
REFERENCE A38146
#authors Diaz, A.; Reginato, A.M.; Jimenez, S.A.
#journal J. Biol. Chem. (1992) 267:10816-10822
#title Alternative splicing of human prostaglandin G/H synthase mRNA
and evidence of differential regulation of the resulting
transcripts by transforming growth factor beta 1,
interleukin 1 beta, and tumor necrosis factor alpha.
#cross-references MUID:92288138
#accession B38146
#molecule_type mRNA
#residues 1-599 #label DIA
#cross-references GB:S36271; NID:g249625; PID:g249626
#experimental_source lung fibroblast
#note sequence extracted from NCBI backbone (NCBIN:103945,
NCBIP:103946)
REFERENCE S50181
#authors Barnett, J.; Chow, J.; Ives, D.; Chiou, M.; Mackenzie, R.;
Osen, E.; Nguyen, B.; Tsing, S.; Bach, C.; Freire, J.;
Chan, H.; Sigal, E.; Ramesha, C.
#journal Biochim. Biophys. Acta (1994) 1209:130-139
#title Purification, characterization and selective inhibition of
human prostaglandin G/H synthase 1 and 2 expressed in the
baculovirus system.
#cross-references MUID:95035046
#accession S50181
#status preliminary
#molecule_type protein
#residues 24-32 #label BAR
COMMENT This enzyme is bifunctional with the fatty acid cyclooxygenase
activity and prostaglandin hydroperoxidase activity.
GENETICS
#gene GDB:PTGS1
#cross-references GDB:128070; OMIM:176805
#map_position 9q32-9q33.3
KEYWORDS alternative splicing; oxidoreductase
FEATURE 1-23
#domain signal sequence #status predicted #label SIG\
#product prostaglandin-endoperoxide synthase #status
experimental #label MAT
24-599 #length 599 #molecular-weight 68656 #checksum 4138
SUMMARY
Query Match 10.8%; Score 101; DB 2; Length 599;
Best Local Similarity 51.9%; Pred. No. 1.39e-01;
Matches 14; Conservative 5; Mismatches 7; Indels 1; Gaps 1;
Db 10 LLFLLLPPLVLLADPGATPVNPPC 36
QY 15 LLLLLLLP-LALLSSGAAQPLPDC 40

RESULT 13
ENTRY T00026 #type complete
TITLE brain-specific angiogenesis inhibitor 1 - human
ALTERNATE_NAMES BA11 protein
ORGANISM #formal_name Homo sapiens #common_name man
DATE 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
12-Feb-1999
ACCESSIONS T00026
REFERENCE Z14064
#authors Nishimori, H.; Shiratsuchi, T.; Urano, T.; Kimura, Y.;
Kiyono, K.; Tatsumi, K.; Yoshida, S.; Ono, M.; Kuwano, M.;

```



\*\*\*\*\*

W P E S R E L H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Aug 20 20:58:15 1999; MasPar time 6.37 Seconds  
239.851 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-938-548B-4  
Description: (1-28) from US08938548B.pep  
Perfect Score: 196  
Sequence: 1 RSGPPGLOGRLQRLQASGNHAAIGLTM 28

Scoring table: PAM 150  
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl9

1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human  
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle  
9:sp.phase 10:sp.plant 11:sp.rodent 12:sp.unclassified  
13:sp.invertebrate 14:sp.virus

Statistics: Mean 30.076; Variance 50.517; scale 0.595

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	196	100.0	131	4	043612 PREPRO-OREXIN.	1.42e-25
2	193	98.5	131	6	077668 PREPRO-OREXIN PRECURSO	6.55e-25
3	190	96.9	130	11	055241 HYPOCRETIN (PREPRO-ORE	3.01e-24
4	190	96.9	130	11	055232 PREPRO-OREXIN.	3.01e-24
5	76	38.8	1174	6	095168 TIGHT JUNCTION PROTEIN	2.25e-01
6	71	36.2	511	10	080740 T13D8.6 PROTEIN.	1.45e+00
7	70	35.7	249	11	P97382 K+ CHANNEL BETA4 SUBUN	2.09e+00
8	70	35.7	1277	11	Q35821 PAR INTERACTING PROTEI	2.09e+00
9	69	35.2	439	2	Q52495 HRPE.	3.00e+00
10	69	35.2	568	5	Q27212 ARTICULIN P60.	3.00e+00
11	69	35.2	1465	5	Q17909 H06001.2 PROTEIN.	3.00e+00
12	68	34.7	162	2	085680 RNA POLYMERASE SIGMA-L	4.29e+00
13	68	34.7	178	2	Q31955 YONC PROTEIN.	4.29e+00
14	68	34.7	178	9	064066 HYPOTHETICAL 19.6 KD P	4.29e+00
15	68	34.7	580	11	Q60811 RETINOID X RECEPTOR IN	4.29e+00
16	67	34.2	145	10	Q48721 T9J22.21 PROTEIN.	6.12e+00
17	67	34.2	161	2	085683 RNA POLYMERASE SIGMA-L	6.12e+00
18	67	34.2	282	2	087877 D-SUBUNIT OF BENZOYL-C	6.12e+00
19	67	34.2	462	2	P95644 RNA POLYMERASE SIGMA F	6.12e+00
20	67	34.2	510	2	Q59913 RNA POLYMERASE SIGMA	6.12e+00

21	67	34.2	511	2	050539 MAJOR VEGETATIVE SIGMA	6.12e+00
22	67	34.2	514	2	P77951 RNA POLYMERASE SIGMA F	6.12e+00
23	67	34.2	525	2	Q59813 RNA POLYMERASE SIGMA F	6.12e+00
24	67	34.2	528	2	Q59814 RNA POLYMERASE SIGMA F	6.12e+00
25	67	34.2	541	2	Q69851 HYPOTHETICAL 60.1 KD P	6.12e+00
26	67	34.2	972	14	P89521 POLYPROTEIN.	6.12e+00
27	67	34.2	972	14	O55651 VP2,NS,VP3.	6.12e+00
28	67	34.2	1420	10	O81016 PUTATIVE ABC TRANSPORT	6.12e+00
29	67	34.2	1841	2	O33958 TYLACTONE SYNTHASE MOD	6.12e+00
30	67	34.2	2115	4	Q14980 NUMA PROTEIN.	6.12e+00
31	67	34.2	2591	2	O54959 PRISTINAMYCIN I SYNTHA	6.12e+00
32	66	33.7	213	2	O50648 TNA2, TNA1, PARTIAL AN	8.69e+00
33	66	33.7	331	13	Q91640 LEUCINE ZIPPER WITH BA	8.69e+00
34	66	33.7	490	11	O54797 STEROID 21-HYDROXYLASE	8.69e+00
35	66	33.7	1163	4	O60312 KIAA0566 PROTEIN (FRAG	8.69e+00
36	66	33.7	1344	11	O35851 P160 MYB-BINDING PROTE	8.69e+00
37	65	33.2	178	11	O62041 MOUSE 57-KD CALCIUM-BI	1.23e+01
38	65	33.2	312	2	O33744 HYPOTHETICAL 33.8 KD P	1.23e+01
39	65	33.2	381	4	Q92931 3-HYDROXYISOBUTYRYL-CO	1.23e+01
40	65	33.2	443	2	O50205 TRANSCRIPTIONAL ACTIVA	1.23e+01
41	65	33.2	457	5	Q19452 F14D7.2 PROTEIN.	1.23e+01
42	65	33.2	481	3	O42651 HYPOTHETICAL 55.5 KD P	1.23e+01
43	65	33.2	973	10	O49541 OXOGLUTARATE DEHYDROGE	1.23e+01
44	64	32.7	263	8	Q31722 ORF263.	1.73e+01
45	64	32.7	1311	4	O14976 HSGAK.	1.73e+01

## ALIGNMENTS

RESULT 1  
ID O43612 PRELIMINARY; PRT; 131 AA.  
AC O43612;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE PREPRO-OREXIN.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
CC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98150861.  
RA SAKURAI T., ANEMIYA A., ISHII M., MATSUZAKI I., CHEMELLI R.M.,  
RA TANAKA H., WILLIAMS S.C., RICHARDSON J.A., KOZLOWSKI G.P., WILSON S.,  
RA ARCH J.R.S., BUCKINGHAM R.E., HAYNES A.C., CARR S.A., ANNAN R.S.,  
RA MCNULTY D.E., LIU W.-S., TERRETT J.A., ELSHOURBAGY N.A., BERGSMAN D.J.,  
RA YANAGISAWA M.;  
RT "Orexins and orexin receptors: a family of hypothalamic neuropeptides  
RT and G protein-coupled receptors that regulate feeding behavior.";  
RL CELL 92:573-585(1998).  
DR EMBL; AF041240; G2897118; -;  
SQ SEQUENCE 131 AA; 13363 MW; 2C11048A CRC32;  
  
Query Match 100.0%; Score 196; DB 4; Length 131;  
Best Local Similarity 100.0%; Pred. No. 1.42e-25;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Db 70 RSGPPGLOGRLQRLQASGNHAAIGLTM 97  
|||||  
Qy 1 RSGPPGLOGRLQRLQASGNHAAIGLTM 28  
  
RESULT 2  
ID O77668 PRELIMINARY; PRT; 131 AA.  
AC O77668;  
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE PREPRO-OREXIN PRECURSOR.  
OS SUS SCROFA (PIG).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
CC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.  
RN [1]

[illegible]



\*\*\*\*\*

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 20:57:47 1999; MasPar time 3.36 Seconds

Tabular output not generated.

233.383 MILLION cell updates/sec

Title: &gt;US-08-938-548B-4

Description: (1-28) from US08938548B.ppt

Perfect Score: 196  
Sequence: 1 RSGPPGLQGR LQRLLOASGNHAA GILTM 28

Scoring table: PAM 150

Gap	15
-----	----

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot37

1:swissprot

Statistics: Mean 31.486; Variance 47.986; scale 0.656

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query		DB	ID	Description	Pred. No.
	Score	Match				
1	71	36.2	1	VE2_HPVE63	REGULATORY PROTEIN E2.	5.54e+01
2	69	35.2	1	ALD_MOUSE	ADRENOLEUKODYSTROPHY P	1.20e+00
3	67	34.2	79	NIFU_FRAAL	NIFU PROTEIN.	2.55e+00
4	67	34.2	442	HRB_STRCO	RNA POLYMERASE PRINCIP	2.55e+00
5	67	34.2	1477	YORL_YEAST	OLIGOMYCIN RESISTANCE	2.55e+00
6	66	33.7	461	VE2_HPVO9	REGULATORY PROTEIN E2.	3.69e+00
7	66	33.7	633	NOQ3_RHLSB	PROBABLE SULFATE ADENY	3.69e+00
8	65	33.2	445	RFLM_HUMAN	MITOCHONDRIAL PEPTIDE	5.33e+00
9	65	33.2	1692	CYAA_SCHPO	ADENYLATE CYCLASE [EC	5.33e+00
10	64	32.7	394	RT04_YEAST	MITOCHONDRIAL 40S RIBO	7.66e+00
11	64	32.7	400	ASSY_SYNY3	ARGININOSUCCINATE SYN	7.66e+00
12	64	32.7	556	YEAL_ECOLI	HYPOTHETICAL 63.2 KD P	7.66e+00
13	64	32.7	690	VTER_EBV	PROBABLE DNA PACKAGING	7.66e+00
14	64	32.7	879	YDHB_ECOLI	HYPOTHETICAL 96.8 KD P	7.66e+00
15	63	32.1	199	IL1L_MOUSE	INTERLEUKIN-11 PRECUR	1.10e+01
16	63	32.1	248	PSPA_HUMAN	PULMONARY SURFACTANT-A	1.10e+01
17	63	32.1	256	YREC_SYNP2	HYPOTHETICAL 28.7 KD P	1.10e+01
18	63	32.1	324	ANX3_RAT	ANNEXIN III (LIPOCORTI	1.10e+01
19	63	32.1	429	UL88_HCMVA	PROTEIN UL88.	1.10e+01
20	63	32.1	878	ECR_DROME	ECDYSONE RECEPTOR	1.10e+01
21	63	32.1	972	POLS_IPNVJ	STRUCTURAL POLYPROTEIN	1.10e+01
22	63	32.1	1157	PEXI_PICPA	PEROXISOME BIOSYNTHESI	1.10e+01
23	63	32.1	1992	TR12_HUMAN	THYROID RECEPTOR INTER	1.10e+01

Query Match 36.2%; Score 71; DB 1; Length 398;  
Best Local Similarity 55.6%; Pred. No. 5.54e-01;

24	62	31.6	187	1	MT28_YEAST	TRANSCRIPTIONAL ACTIVA	1.56e+01
25	62	31.6	374	1	CGSQ_BOVIN	RETINA SPECIFIC REGULA	1.56e+01
26	62	31.6	511	1	RTAQ_THEAQ	THERMOSTABLE CARBOXYE	1.56e+01
27	62	31.6	624	1	SIRB_SYPN7	SULFITE REDUCTASE (FER	1.56e+01
28	62	31.6	928	1	VGLB_MCMVS	GLYCOPROTEIN B PRECURS	1.56e+01
29	62	31.6	972	1	POLS_IPNVN	STRUCTURAL POLYPEPTIDE	1.56e+01
30	62	31.6	1176	1	NIR_NEUCR	NITRITE REDUCTASE (NAD	1.56e+01
31	62	31.6	1690	1	KFIA_HUMAN	KINESIN-LIKE PROTEIN K	1.56e+01
32	62	31.6	1695	1	KFIA_MOUSE	KINESIN-LIKE PROTEIN K	1.56e+01
33	62	31.6	2504	1	FAS_HUMAN	FATTY ACID SYNTHASE (E	1.56e+01
34	61	31.1	180	1	NEF_HV2NZ	NEGATIVE FACTOR (F-PRO	2.22e+01
35	61	31.1	249	1	PSPA_PIG	PULMONARY SURFACTANT-A	2.22e+01
36	61	31.1	410	1	B7_USTMA	MATING-TYPE LOCUS ALLE	2.22e+01
37	61	31.1	520	1	YF77_YEAST	HYPOPHETICAL 57.9 KD P	2.22e+01
38	61	31.1	799	1	AFSK_SRRCO	SERINE/THREONINE PROTE	2.22e+01
39	60	30.6	274	1	DCOP_MYCTO	OTIDINE 5'-PHOSPHATE	3.13e+01
40	60	30.6	302	1	YP95_CAEEL	HYPOHETICAL 34.8 KD P	3.13e+01
41	60	30.6	367	1	YOPM_YERPE	OUTER MEMBRANE PROTEIN	3.13e+01
42	60	30.6	380	1	F812_MOUSE	FACTOR VIII INTRON 22	3.13e+01
43	60	30.6	702	1	YCBY_ECOLI	HYPOPHETICAL 78.9 KD P	3.13e+01
44	60	30.6	780	1	ACON_BOVIN	ACONITATE HYDRATASE, M	3.13e+01
45	60	30.6	5217	1	HTS1_COCCA	HC-TOXIN SYNTHETASE, (	3.13e+01

## ALIGNMENTS

```

RESULT 1
ID VQ22_HPV63 STANDARD; PRT; 398 AA.
AC Q07850;
DT DT
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE REGULATORY PROTEIN E2.
DE E2.
GN HUMAN PAPILLOMAVIRUS TYPE 63.
OS VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPAPOVIRIDAE; PAPILLOMAVIRUS.
OC [1]
RN SEQUENCE FROM N.A.
RP MEDLINE: 93276568.
RX EGAWA K., DELIUS H., MATSUKURA T., KAWASHIMA M., DE VILLIERS E.M.;
RT "Two novel types of human papillomavirus, HPV 63 and HPV 65:
RT comparisons of their clinical and histological features and DNA
RT companions to other HPV types.";
RT VIROLOGY 194:789-799(1993).
RL
RL -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCCNNNNNGGT-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SUBUNIT: BINDS DNA AS A DIMER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X70828; G312096; -.
CC DR PFAM; PF00508; E2_N; 1.
CC DR PFAM; PF00511; E2_C; 1.
CC DR HSP; P17383; 1DHM.
CC DR EARLY PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING;
CC TRANS-ACTING FACTOR; DNA REPLICATION; REPRESSOR; NUCLEAR PROTEIN.
SQ SEQUENCE 398 AA; 45450 MW; C9B80CE0 CRC32;

```

Query Match  
Best Local Similarity 31.8%; Score 67; DB 1; Length 442;  
Matches 7; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

Db 370 LQEQHSLVDTLSERAGVVSM 391  
|||:::||||:  
Y 7 LQGRLLQLQASGNHAAGILTM 28

Query Match  
Best Local Similarity 34.2%; Score 67; DB 1; Length 1477;  
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 499 GRLOSLEAPDDPNQMIEM 518  
|||||:::|  
Y 9 GRLLRLLQASGNHAAGILTM 28

Query Match  
Best Local Similarity 33.7%; Score 66; DB 1; Length 461;  
Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Db 358 RSVGAGHHGLRLALAEAKDPPLMLL 383  
|||:::||||:  
Y 1 RSGPFLQGRLLRLLQASGNHAAGIL 26

Query Match 34.2%; Score 67; DB 1; Length 442;  
Best Local Similarity 31.8%; Pred. No. 2.55e+00;  
Matches 7; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

Db 370 LQEQHSLVDTLSERAGVVSM 391  
|||:::||||:  
Qy 7 LQGRLLQLQASGNHAAGILTM 28

RESULT 5  
ID YORL\_YEAST STANDARD; PRT; 1477 AA.  
AC P3049;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE OLIGOMYCIN RESISTANCE ATP-DEPENDENT PERMEASE YORL.  
GN YORL OR YGR281W.  
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;  
OC SACCHAROMYCETACEAE; SACCHAROMYCES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 96069397.  
RA KATZMANN D.J., HALLSTROM T.C., VOET M., WYSOCK W., GOLIN J.,  
RA VOLCKAERT G., MOYE-ROWLEY W.S.;  
RT "Expression of an ATP-binding cassette transporter-encoding gene  
RT (YORL) is required for oligomycin resistance in Saccharomyces  
RT cerevisiae.";  
RT MOL. CELL. BIOL. 15:6875-6883(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C / FY1679;  
RX MEDLINE; 97245295.  
RA VOLCKAERT G., VOET M., ROBBEN J.;  
RT "Sequence analysis of a near-subtelomeric 35.4 kb DNA segment on the  
RT right arm of chromosome VII from Saccharomyces cerevisiae carrying  
RT the MAL1 locus reveals 15 complete open reading frames, including  
RT ZUO1, BGL2 and BIO2 genes and an ABC transporter gene.";  
RL YEAST 13:251-259(1997).  
CC -!- FUNCTION: REQUIRED FOR OLIGOMYCIN RESISTANCE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
CC (ABC TRANSPORTERS). BELONGS TO THE MDR SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
CC EMBL; Z73066; E243252; -  
CC SGD; L0003083; YORL.  
DR DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
DR PFAM; PF00005; ABC\_tran; 2.  
DR PFAM; PF00664; ABC\_membrane; 2.  
DR HSSP; P13569; INDB.  
KW ATP-BINDING; TRANSMEMBRANE; GLYCOPROTEIN; TRANSPORT.  
FT TRANSMEM 207 227 POTENTIAL.  
FT TRANSMEM 250 270 POTENTIAL.  
FT TRANSMEM 329 349 POTENTIAL.  
FT TRANSMEM 350 370 POTENTIAL.  
FT TRANSMEM 434 454 POTENTIAL.  
FT TRANSMEM 479 499 POTENTIAL.  
FT TRANSMEM 616 636 POTENTIAL.  
FT TRANSMEM 893 913 POTENTIAL.  
FT TRANSMEM 941 961 POTENTIAL.  
FT TRANSMEM 1028 1048 POTENTIAL.  
FT TRANSMEM 1118 1138 POTENTIAL.  
FT TRANSMEM 1142 1162 POTENTIAL.

Query Match 34.2%; Score 67; DB 1; Length 442;  
Best Local Similarity 31.8%; Pred. No. 2.55e+00;  
Matches 7; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

Db 370 LQEQHSLVDTLSERAGVVSM 391  
|||:::||||:  
Qy 7 LQGRLLQLQASGNHAAGILTM 28

RESULT 5  
ID YORL\_YEAST STANDARD; PRT; 1477 AA.  
AC P3049;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE OLIGOMYCIN RESISTANCE ATP-DEPENDENT PERMEASE YORL.  
GN YORL OR YGR281W.  
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;  
OC SACCHAROMYCETACEAE; SACCHAROMYCES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 96069397.  
RA KATZMANN D.J., HALLSTROM T.C., VOET M., WYSOCK W., GOLIN J.,  
RA VOLCKAERT G., MOYE-ROWLEY W.S.;  
RT "Expression of an ATP-binding cassette transporter-encoding gene  
RT (YORL) is required for oligomycin resistance in Saccharomyces  
RT cerevisiae.";  
RT MOL. CELL. BIOL. 15:6875-6883(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C / FY1679;  
RX MEDLINE; 97245295.  
RA VOLCKAERT G., VOET M., ROBBEN J.;  
RT "Sequence analysis of a near-subtelomeric 35.4 kb DNA segment on the  
RT right arm of chromosome VII from Saccharomyces cerevisiae carrying  
RT the MAL1 locus reveals 15 complete open reading frames, including  
RT ZUO1, BGL2 and BIO2 genes and an ABC transporter gene.";  
RL YEAST 13:251-259(1997).  
CC -!- FUNCTION: REQUIRED FOR OLIGOMYCIN RESISTANCE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
CC (ABC TRANSPORTERS). BELONGS TO THE MDR SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
CC EMBL; Z73066; E243252; -  
CC SGD; L0003083; YORL.  
DR DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
DR PFAM; PF00005; ABC\_tran; 2.  
DR PFAM; PF00664; ABC\_membrane; 2.  
DR HSSP; P13569; INDB.  
KW ATP-BINDING; TRANSMEMBRANE; GLYCOPROTEIN; TRANSPORT.  
FT TRANSMEM 207 227 POTENTIAL.  
FT TRANSMEM 250 270 POTENTIAL.  
FT TRANSMEM 329 349 POTENTIAL.  
FT TRANSMEM 350 370 POTENTIAL.  
FT TRANSMEM 434 454 POTENTIAL.  
FT TRANSMEM 479 499 POTENTIAL.  
FT TRANSMEM 616 636 POTENTIAL.  
FT TRANSMEM 893 913 POTENTIAL.  
FT TRANSMEM 941 961 POTENTIAL.  
FT TRANSMEM 1028 1048 POTENTIAL.  
FT TRANSMEM 1118 1138 POTENTIAL.  
FT TRANSMEM 1142 1162 POTENTIAL.

Query Match 34.2%; Score 67; DB 1; Length 1477;  
Best Local Similarity 40.0%; Pred. No. 2.55e+00;  
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 499 GRLOSILLEAPDDPNQMIEM 518  
|||||:::|:  
Qy 9 GRLLRLLQASGNHAAGILTM 28

RESULT 6  
ID VE2\_HPV09 STANDARD; PRT; 461 AA.  
AC P36780;  
DT 01-JUN-1994 (REL. 29, CREATED)  
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE REGULATORY PROTEIN E2.  
GN E2.  
OS HUMAN PAPILLOMAVIRUS TYPE 9.  
OC VIRUSES; DSNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; PAPILLOMAVIRUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX DELIUS H., HOFMANN B.;  
RA "Primer-directed sequencing of human papillomavirus types.";  
RL CURR. TOP. MICROBIOL. IMMUNOL. 186:13-31(1994).  
CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.  
CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCNNNNNGGT-3') PRESENT  
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER  
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING ON E2RE'S POSITION  
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS  
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION  
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA  
CC REPLICATION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- SUBUNIT: BINDS DNA AS A DIMER.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
CC EMBL; X74464; G397073; -  
CC PIR; S36593; S36593.  
DR PFAM; PF00508; E2\_N; 1.  
DR PFAM; PF00511; E2\_C; 1.  
DR HSSP; P17383; IDHM.  
KW EARLY PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING;  
KW TRANS-ACTING FACTOR; DNA REPLICATION; REPRESSOR; NUCLEAR PROTEIN.  
SQ SEQUENCE 461 AA; 52141 MW; DA2B4125 CRC32;

Query Match 33.7%; Score 66; DB 1; Length 461;  
Best Local Similarity 38.5%; Pred. No. 3.69e+00;  
Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Db 358 RSVGAGHHGRLARLAIEAKDPPLMLL 383  
|||:::||||:  
Qy 1 RSGPPGLQGRLLRLLQASGNHAAGIL 26

SEQUENCE FROM N. A.  
RP STRAIN-S286C AB972;  
RC MEDLINE; 94376003;  
EX  
RA JONSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,  
RA DU Z., FAVELLO A., FULTON L., GATTONG S., GEISEL C., KIRSTEN J.,  
RA KUCHARA I., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,  
RA KATREBBA T., LOUIS E. J., MACRI C., MARDIS E., MENEZES S., MOUSER L.,  
RA MURPHY P.

RA NIAN M., KIFANI L., KILS L., SI PETER H., TREVASKIS E., VAUGHAN K.,  
RA VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,  
RA VAUDIN M.;  
RT "complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome  
RT VIII";  
CC SCIENCE 265:2077-2082(1994).  
CC -!- FUNCTION: COMPONENT OF THE SMALL SUBUNIT OF MITOCHONDRIAL  
CC PEROXISOME.

CC -! SUBCELLULAR LOCATION: MITOCHONDRIAL.  
CC -! SIMILARITY: BELONGS TO THE S2P FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC -----  
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute.

CC use by non-profit institutions as long as its content is in no way  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC the European Bioinformatics Institute. Usage by and for commercial  
CC modified and this statement is not removed.

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC -----  
CC -----  
CC -----

C		EMBL; M82841; GI71982; -;
DR		EMBL; U10555; G500819; -;

DR DR  
PIR; S27429; S27429.  
PIR; A42115; A42115.  
DR DR  
SGD; L0001154; MRP4.  
DR DR

DR	PROSITE; PS00962; RIBOSOMAL_S2_1; 1.
DR	PROSITE; PS00963; RIBOSOMAL_S2_2; 1.
DR	PROSITE; PS00963; RIBOSOMAL_S2_2; 1.
DR	PROSITE; PS00963; RIBOSOMAL_S2_2; 1.

DK PFAM; PF00318; S2; 1.  
RIBOSOMAL PROTEIN; MITOCHONDRION.  
SQ SEQUENCE 394 AA; 44151 MW; 5226C1B0 CRC32;

Query Match	32.7%	Score 64;	DB 1;	Length 394;
Best Local Similarity	61.5%	Prod No	7	566400;

Best local similarity 0.75, freq. NO. 7.00E+00,  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 378 RAGQRGLQNRLR 390  
 | : | | | | |  
 Qv 1 RSGPPGLQRIOR 13

[illegible]

RESULT	II	STANDARD;	PRT;	400 AA.
ID	ASSY_SYNY3			
AC	P77973;			

DT	01-NOV-1997	(REL.	35,	CREATED)
DT	01-NOV-1997	(REL.	35,	LAST SEQUENCE UPDATE)
DT	01-NOV-1997	(REL.	35,	LAST SEQUENCE UPDATE)
DT	01-NOV-1997	(REL.	35,	LAST ANNOTATION UPDATE)

01-NOV-1997 (REL: 33, LAST AMENDMENT UPDATE)  
DE ARGININOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPARTATE  
DE LIGASE).

GN ARGG OR SLR0585.  
OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).  
OC BACTERIA: CYANOBACTERIA: CHROOCOCCALES. SYNECHOCYSTIS

```
RN [1]
RP SEQUENCE FROM N.A.
RE NUMBER OF REPLICATES
RW WEIGHTING FACTOR
```

RX  
 MEDLINE: 9/061201.  
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,  
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.  
 RA

RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,  
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,  
DA

"Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the

RT entire genome and assignment of potential protein-coding regions.";  
RL DNA RES. 3:109-136(1996).  
CC -1- CATALYTIC ACTIVITY. ATP + L-CITRULLINE + L-ASPARTATE -> AMP +

CC PYROPHOSPHATE + L-ARGININOSUCCINATE.  
CC -1- PATHWAY: UREA CYCLE, PENULTIMATE STEP OF THE ARGININE BIOSYNTHETIC

CC  
PATHWAY.







#submission submitted to the EMBL Data Library, April 1998  
#description Arabidopsis thaliana chromosome II BAC T9J22 genomic  
#accession T00987  
##status preliminary; translated from GB/EMBL/DBDJ  
##molecule\_type DNA  
##residues 1-145 ##label ROU  
##cross-references EMBL:AC002505; NID:g2739379

GENETICS

#map\_position II  
#introns 17/1; 49/3; 78/2; 123/2  
#note T9J22.21

SUMMARY

Query Match 34.2%; Score 67; DB 2; Length 145;  
Best Local Similarity 40.0%; Pred. No. 6.46e+00;  
Matches 8; Conservative 8; Mismatches 4; Indels 0; Gaps 0;  
Db 119 PGLEGWVESIMEALGAHADS 138  
QY 5 PGLOQLRLQASGNHAAG 24

RESULT 6  
ENTRY S11712 #type complete  
TITLE transcription initiation factor sigma hrdB - Streptomyces  
ALTERNATE\_NAMES coelicolor  
ORGANISM sigma-like transcription factor  
DATE 03-May-1994 #sequence\_revision 20-Feb-1995 #text\_change

ACCESSIONS S11712; JH0496; B40116  
REFERENCE S11711  
#authors Tanaka, K.; Shiina, T.; Takahashi, H.  
#submission submitted to the EMBL Data Library, May 1990  
#description Multiple homolog genes for principal sigma subunit of  
Streptomyces coelicolor A3(2).

#accession S11712  
##molecule\_type DNA  
##residues 1-442 ##label EMB  
##cross-references EMBL:X52983; NID:g48744; PID:g48745  
##experimental\_source strain A3(2)

REFERENCE JH0496  
#authors Shiina, T.; Tanaka, K.; Takahashi, H.  
#journal Gene (1991) 107:145-148  
#title Sequence of hrdB, an essential gene encoding sigma-like  
transcription factor of Streptomyces coelicolor A3(2):  
homology to principal sigma factors.

#cross-references MUID:92077425  
#contents A3[2]  
#accession JH0496

##molecule\_type DNA  
##residues 1-429, 'OG', 432-442 ##label SHI  
##cross-references EMBL:X52983  
##note the authors translated the codon CAG for residue 430 as  
His and GGG for residue 431 as Pro

REFERENCE A40116  
#authors Tanaka, K.; Shiina, T.; Takahashi, H.  
#journal Science (1988) 242:1040-1042  
#title Multiple principal sigma factor homologs in eubacteria:  
identification of the "rpoD box".

#cross-references MUID:89058616  
#accession B40116

##status preliminary; nucleic acid sequence not shown; not  
compared with conceptual translation

##molecule\_type DNA  
##residues 237-288 ##label TAN  
##note This protein is the functional homolog of the principal sigma  
factors involved in the transcription of housekeeping genes.

GENETICS

#gene hrdB  
CLASSIFICATION #superfamily Streptomyces transcription initiation factor  
sigma; transcription initiation factor sigma katF homology

KEYWORDS DNA binding; sigma factor; transcription initiation  
FEATURE  
211-437 #domain transcription initiation factor sigma katF  
homology #label KTF

SUMMARY

Query Match 34.2%; Score 67; DB 2; Length 442;  
Best Local Similarity 31.8%; Pred. No. 6.46e+00;  
Matches 7; Conservative 11; Mismatches 4; Indels 0; Gaps 0;  
Db 370 LQEOHLHSLDLSEREAGVWSM 391  
QY 7 LQRLRLQASGNHAAGILTM 28

RESULT 7  
ENTRY S41307 #type complete  
TITLE transcription initiation factor sigma - Streptomyces griseus  
ORGANISM #formal\_name Streptomyces griseus  
DATE 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change

ACCESSIONS S41307  
REFERENCE S41306  
#authors Marcos, A.T.; Diez, B.; Gutierrez, S.; Fernandez, F.J.;  
Velasco, J.; Martin, J.F.  
#submission submitted to the EMBL Data Library, December 1993  
#description Organization and expression of the hrdB-sprC gene cluster of  
streptomyces griseus encoding a sigma factor protein and a  
serine protease. Role on growth and sporulation of  
streptomyces.

#accession S41307  
##status preliminary  
##molecule\_type DNA  
##residues 1-510 ##label MAR  
##cross-references EMBL:X75952; NID:g440164; PID:g581664

GENETICS

#start\_codon GTG  
CLASSIFICATION #superfamily Streptomyces transcription initiation factor  
sigma; transcription initiation factor: sigma katF homology

KEYWORDS DNA binding; sigma factor; transcription initiation  
FEATURE  
279-505 #domain transcription initiation factor sigma katF  
homology #label KTF

SUMMARY

Query Match 34.2%; Score 67; DB 2; Length 510;  
Best Local Similarity 31.8%; Pred. No. 6.46e+00;  
Matches 7; Conservative 11; Mismatches 4; Indels 0; Gaps 0;  
Db 438 LQEOHLHSLDLSEREAGVWSM 459  
QY 7 LQRLRLQASGNHAAGILTM 28

RESULT 8  
ENTRY JN0443 #type complete  
TITLE transcription initiation factor sigma homolog hrdB -  
Streptomyces aureofaciens

ORGANISM #formal\_name Streptomyces aureofaciens  
DATE 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change

ACCESSIONS JN0443  
REFERENCE JN0442

#authors Kormanec, J.; Farkasovsky, M.; Poutuckova, L.  
#journal Gene (1992) 122:63-70  
#title Four genes in Streptomyces aureofaciens containing a domain  
characteristic of principal sigma factors.

#cross-references MUID:93083996  
#accession JN0443  
##molecule\_type DNA  
##residues 1-525 ##label KOR  
##cross-references GB:M90411; NID:g153305; PID:g153306

GENETICS

```

TITLE      NuMA protein - human
ORGANISM   #formal_name Homo sapiens #common_name man
DATE       19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change
          10-Sep-1997
ACCESSIONS S23647
REFERENCE   S23647
#authors   Yang, C.H.; Lambie, E.J.; Snyder, M.
#journal   J. Cell Biol. (1992) 116:1303-1317
#title     NuMA: an unusually long coiled-coil related protein in the
          mammalian nucleus.
#cross-references EMBL:Z11583; NID:g35118; PID:g35119
          the authors translated the codon GAG for residue 781 as
          Gly, TTC for residue 1775 as Pro, and GGA for residue
          2067 as Glu
#length 2115 #molecular-weight 238273 #checksum 4391
SUMMARY
Query Match      34.2%; Score 67; DB 2; Length 2115;
Best Local Similarity 52.9%; Pred. No. 6.46e+00;
Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 761 RAGKGLRLQQLGEA 777
   || |||||:|
   1 RSGPPGLQRLQLLQA 17
   || |||||:|

RESULT 13
ENTRY   S36593 #type complete
TITLE   E2 protein - human papillomavirus type 9
ORGANISM #formal_name human papillomavirus type 9
DATE     20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
          08-Sep-1997
ACCESSIONS S36593
REFERENCE   S36469
#authors   Delius, H.; Hofmann, B.
#description submitted to the EMBL Data Library, August 1993
#accession S36593
#molecule_type DNA
#residues 1-461 #label DEL
#cross-references EMBL:X74464; NID:g397068; PID:g397073
CLASSIFICATION #superfamily papillomavirus E2 protein
KEYWORDS DNA binding; early protein; transcription regulation
SUMMARY #length 461 #molecular-weight 52141 #checksum 478

Query Match      33.7%; Score 66; DB 2; Length 461;
Best Local Similarity 38.5%; Pred. No. 9.07e+00;
Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Db 358 RSVGAGHGRRLRLAEAKDPPLMLL 383
   || |||||:|
   1 RSGPPGLQRLQLLQA 26
   || |||||:|

RESULT 14
ENTRY   S26481 #type complete
TITLE   calcium-binding protein, 57K - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE     06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
          10-Sep-1997
ACCESSIONS S26481
REFERENCE   S26481
#authors   Tuan, R.S.; Fitzpatrick, D.
#submission submitted to the EMBL Data Library, September 1990
#description Structural analysis of mouse placental 57-KD Calcium-binding
          protein.
#accession S26481
#status preliminary
#molecule_type mRNA

```

```

#residues 1-178 #label TUA
#cross-references EMBL:X56603; NID:g53597; PID:g53598
KEYWORDS calcium binding
SUMMARY #length 178 #molecular-weight 19960 #checksum 5316

Query Match      33.2%; Score 65; DB 2; Length 178;
Best Local Similarity 30.8%; Pred. No. 1.27e+01;
Matches 8; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

Db 147 RGEALGRHLFLHRLRGQKGVV 172
   || |||||:|
   1 RSGPPGLQRLQLLQA 26
   || |||||:|

RESULT 15
ENTRY   A33988 #type complete
TITLE   adenylate cyclase (EC 4.6.1.1) - fission yeast
          (Schizosaccharomyces pombe)
ORGANISM #formal_name Schizosaccharomyces pombe
DATE     23-Mar-1990 #sequence_revision 23-Mar-1990 #text_change
          09-Sep-1997
ACCESSIONS A33988; A33539
REFERENCE   A33988
#authors   Young, D.; Riggs, M.; Field, J.; Vojtek, A.; Broek, D.;
          Wigler, M.
#journal   Proc. Natl. Acad. Sci. U.S.A. (1989) 86:7989-7993
#title     The adenyl cyclase gene from Schizosaccharomyces pombe.
#cross-references MUID:90046723
#accession A33988
#molecule_type DNA
#residues 1-1692 #label YOU
#cross-references GB:M26699; NID:gl73338; PID:gl73339
REFERENCE   A33539
#authors   Yanawaki-Kataoka, Y.; Tamaoki, T.; Choe, H.R.; Tanaka, H.;
          Kataoka, T.
#journal   Proc. Natl. Acad. Sci. U.S.A. (1989) 86:5693-5697
#title     Adenylate cyclases in yeast: a comparison of the genes from
          Schizosaccharomyces pombe and Saccharomyces cerevisiae.
#cross-references MUID:89345533
#accession A33539
#status preliminary
#molecule_type DNA
#residues 1-1692 #label YAM
#cross-references GB:M24942; NID:gl73378; PID:gl73379
#note      the authors translated the codon TGC for residue 626 as
          Ser, and GCC for residue 1243 as Gly
CLASSIFICATION #superfamily leucine-rich alpha-2-glycoprotein repeat
          homology; yeast adenylate cyclase catalytic domain homology
          phosphorus-oxygen lyase
KEYWORDS
FEATURE 1328-1413 #domain yeast adenylate cyclase catalytic domain
          homology #label YACC
SUMMARY #length 1692 #molecular-weight 190332 #checksum 2609

Query Match      33.2%; Score 65; DB 2; Length 1692;
Best Local Similarity 57.1%; Pred. No. 1.27e+01;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 1539 PVQLQGRRLRLIKS 1552
   || |||||:|
   4 PPGGLQRLQLLQA 17
   || |||||:|

Search completed: Fri Aug 20 20:57:30 1999
Job time : 22 secs.

```

\*\*\*\*\*

W P E R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Aug 20 20:59:00 1999; MasPar time 1.72 Seconds  
Tabular output not generated. 165,470 Million cell updates/sec

Title: >US-08-938-548B-4  
Description: (1-28) from US08938548B.pep  
Perfect Score: 196  
Sequence: 1 RSGPGLQRLQLQASGNHAAAILTM 28

Scoring table: PAM 150  
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 20.494; Variance 80.948; scale 0.253

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	68	34.7	716	3	PCT-US95-1 Sequence 4, Applicatio	2.13e+01
2	67	34.2	442	2	US-08-363- Sequence 4, Applicatio	2.59e+01
3	67	34.2	442	2	US-08-363- Sequence 11, Applicati	2.59e+01
4	67	34.2	445	2	US-08-363- Sequence 12, Applicati	2.59e+01
5	67	34.2	445	2	US-08-363- Sequence 5, Applicatio	2.59e+01
6	67	34.2	1841	2	US-08-804- Sequence 6, Applicatio	2.59e+01
7	66	33.7	10	3	PCT-US91-0 Sequence 3, Applicatio	3.14e+01
8	65	33.2	381	2	US-08-858- Sequence 3, Applicatio	3.81e+01
9	65	33.2	381	2	US-08-858- Sequence 1, Applicatio	3.81e+01
10	62	31.6	176	3	PCT-US96-1 Sequence 6, Applicatio	6.78e+01
11	62	31.6	176	2	US-08-659- Sequence 6, Applicatio	6.78e+01
12	62	31.6	2509	1	US-08-469- Sequence 10, Applicati	6.78e+01
13	60	30.6	199	1	US-07-941- Sequence 2, Applicatio	9.90e+01
14	60	30.6	199	1	US-08-792- Sequence 8, Applicatio	9.90e+01
15	60	30.6	199	1	US-07-949- Sequence 4, Applicatio	9.90e+01
16	60	30.6	199	2	US-08-814- Sequence 4, Applicatio	9.90e+01
17	60	30.6	199	1	US-08-115- Sequence 2, Applicatio	9.90e+01
18	60	30.6	199	1	US-08-017- Sequence 2, Applicatio	9.90e+01
19	60	30.6	199	3	PCT-US93-0 Sequence 2, Applicatio	9.90e+01
20	60	30.6	296	1	US-07-745- Sequence 14, Applicati	9.90e+01
21	60	30.6	296	1	US-08-115- Sequence 4, Applicatio	9.90e+01
22	60	30.6	296	1	US-07-941- Sequence 4, Applicatio	9.90e+01
23	60	30.6	296	1	US-08-165- Sequence 14, Applicati	9.90e+01

24	60	30.6	296	1	US-07-921- Sequence 14, Applicati	9.90e+01
25	60	30.6	296	3	PCT-US93-0 Sequence 4, Applicatio	9.90e+01
26	60	30.6	296	3	PCT-US94-1 Sequence 14, Applicati	9.90e+01
27	60	30.6	300	2	US-08-794- Sequence 2, Applicatio	9.90e+01
28	60	30.6	528	2	US-08-363- Sequence 14, Applicati	9.90e+01
29	60	30.6	530	2	US-08-363- Sequence 10, Applicati	9.90e+01
30	60	30.6	530	2	US-08-363- Sequence 2, Applicatio	9.90e+01
31	60	30.6	530	2	US-08-363- Sequence 9, Applicatio	9.90e+01
32	60	30.6	530	2	US-08-363- Sequence 8, Applicatio	9.90e+01
33	60	30.6	530	2	US-08-363- Sequence 3, Applicatio	9.90e+01
34	60	30.6	1271	3	PCT-US94-0 Sequence 2, Applicatio	9.90e+01
35	60	30.6	1271	2	US-08-444- Sequence 2, Applicatio	9.90e+01
36	60	30.6	1271	2	US-08-095- Sequence 2, Applicatio	9.90e+01
37	60	30.6	3724	2	US-08-804- Sequence 10, Applicati	9.90e+01
38	59	30.1	829	1	US-08-446- Sequence 2, Applicatio	1.19e+02
39	59	30.1	829	1	US-08-446- Sequence 2, Applicatio	1.19e+02
40	59	30.1	829	1	US-08-445- Sequence 2, Applicatio	1.19e+02
41	59	30.1	829	1	US-07-670- Sequence 2, Applicatio	1.19e+02
42	59	30.1	829	1	US-08-220- Sequence 2, Applicatio	1.19e+02
43	58	29.6	199	4	5215895-4 Patent No. 5215895.	1.44e+02
44	58	29.6	199	4	5215895-3 Patent No. 5215895.	1.44e+02
45	58	29.6	199	1	US-08-017- Sequence 2, Applicatio	1.44e+02

#### ALIGNMENTS

RESULT 1 PCT-US95-16311-4 STANDARD; PRT; 716 AA.  
XX xxxxxx

Sequence 4, Application PC/TUS9516311

Sequence 4, Application PC/TUS9516311  
GENERAL INFORMATION:

APPLICANT: Moore, David  
APPLICANT: Seol, Wongi  
APPLICANT: Choi, Hwang-Sik  
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING  
POLYPEPTIDES AND RELATED MOLECULES AND METHODS  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street, Suite 3100  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/16311  
FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/372,652

FILING DATE: 13-JAN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul T.

REGISTRATION NUMBER: 30,162

REFERENCE/DOCKET NUMBER: 00786/246001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 716 amino acids

TYPE: amino acid

```

XX Sequence 12, Application US/08363255
CC Patent No. 5783386
CC GENERAL INFORMATION:
CC APPLICANT: JACOBS, JR., WILLIAM R.
CC APPLICANT: BLOOM, BARRY R.
CC APPLICANT: COLLINS, DESMOND M.
CC APPLICANT: de LISLE, GEOFFREY W.
CC APPLICANT: PASCOPELLA, LISA
CC APPLICANT: KAWAKAMI, RIKU P.
CC TITLE OF INVENTION: MYCOBACTERIA VIRULENCE FACTORS AND A
CC TITLE OF INVENTION: NOVEL METHOD FOR THEIR IDENTIFICATION
CC NUMBER OF SEQUENCES: 14
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORRISON & FOERSTER
CC STREET: 755 Page Mill Road
CC CITY: Palo Alto
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94304-1018
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC FILING DATE: 23-DEC-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MONROY, GLADYS H.
CC REGISTRATION NUMBER: 32,430
CC REFERENCE/DOCKET NUMBER: 25237-20002.22
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 813-5600
CC TELEFAX: (415) 494-0792
CC TELEX: 706141
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 445 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC
CC SQ SEQUENCE 445 AA; 48583 MW; 967304 CN;

Query Match 34.2%; Score 67; DB 2; Length 445;
Best Local Similarity 31.8%; Pred. No. 2.59e+01;
Matches 7; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

Db 373 LQEQHLSVLDLTLEREAGVWSM 394
|| :|: :|: :|: :|: :|
Qy 7 LQRLRLQASGNHAGILTM 28

RESULT 5
ID US-08-363-255-5 STANDARD; PRT; 445 AA.
XX
XX
AC xxxxxx
XX
DT
DT
XX
DE
XX
Sequence 5, Application US/08363255
Sequence 5, Application US/08363255
Patent No. 5783386
GENERAL INFORMATION:
APPLICANT: JACOBS, JR., WILLIAM R.
APPLICANT: BLOOM, BARRY R.
APPLICANT: COLLINS, DESMOND M.
APPLICANT: de LISLE, GEOFFREY W.
APPLICANT: PASCOPELLA, LISA
APPLICANT: KAWAKAMI, RIKU P.
TITLE OF INVENTION: MYCOBACTERIA VIRULENCE FACTORS AND A
TITLE OF INVENTION: NOVEL METHOD FOR THEIR IDENTIFICATION
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 23-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, GLADYS H.
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 25237-20002.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 445 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

```

```

CC TITLE OF INVENTION: NOVEL METHOD FOR THEIR IDENTIFICATION
CC NUMBER OF SEQUENCES: 14
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORRISON & FOERSTER
CC STREET: 755 Page Mill Road
CC CITY: Palo Alto
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94304-1018
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC FILING DATE: 23-DEC-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MONROY, GLADYS H.
CC REGISTRATION NUMBER: 32,430
CC REFERENCE/DOCKET NUMBER: 25237-20002.22
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 813-5600
CC TELEFAX: (415) 494-0792
CC TELEX: 706141
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 445 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC
CC SQ SEQUENCE 445 AA; 48583 MW; 967304 CN;

Query Match 34.2%; Score 67; DB 2; Length 445;
Best Local Similarity 31.8%; Pred. No. 2.59e+01;
Matches 7; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

Db 373 LQEQHLSVLDLTLEREAGVWSM 394
|| :|: :|: :|: :|: :|
Qy 7 LQRLRLQASGNHAGILTM 28

RESULT 6
ID US-08-804-227C-6 STANDARD; PRT; 1841 AA.
XX
XX
AC xxxxxx
XX
DT
DT
XX
DE
XX
Sequence 6, Application US/08804227C
Sequence 6, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: DeHoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only

```

PCT-US96-11445-6	STANDARD;	PRT:	176 AA.
xxxxxx			
Sequence 6, Application PC/TUS9611445			
Sequence 6, Application PC/TUS9611445			
GENERAL INFORMATION:			
APPLICANT: The Regents of the University of California			
TITLE OF INVENTION: Isolation of Novel HIV-2 Proviruses			
NUMBER OF SEQUENCES: 50			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Robbins, Berliner & Carson			
STREET: 201 N. Figueroa Street, 5th Floor			
CITY: Los Angeles			
STATE: California			
COUNTRY: USA			
ZIP: 90012-2628			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: PatentIn Release #1.0, Version #1.30			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: PCT/US96/11445			
FILING DATE:			
CLASSIFICATION:			
ATTORNEY/AGENT INFORMATION:			
NAME: Berliner, Robert			
REGISTRATION NUMBER: 20,121			
REFERENCE/DOCKET NUMBER: 5555-399C1			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: (213) 977-1001			
TELEFAX: (213) 977-1003			
INFORMATION FOR SEQ ID NO: 6:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 176 amino acids			
TYPE: amino acid			
STRANDEDNESS:			
TOPOLOGY: linear			
MOLECULE TYPE: protein			
FEATURE:			
NAME/KEY: Protein			
LOCATION: 1..176			
OTHER INFORMATION: /note= "rev protein encoded by HIV-2KR"			
SEQUENCE 176 AA; 19639 MW; 145168 CN;			
Query Match 31.6%; Score 62; DB 3; Length 176;			
Best Local Similarity 40.0%; Pred. No. 6.78e+01;			
Matches: 10; Conservative 5; Mismatches 9; Indels 1; Gaps 1;			
Dd 4 RADEEGLQG-LIRLLHQTDYPQGL 27			
I:       : :			
Oy 1 RSGPPGLQGRLLRLLQASGNAAGI 25			
RESULT 11			
ID US-08-659-251-6 STANDARD;			PRT: 176 AA.
XXXXXX			
Sequence 6, Application US/08659251			
Sequence 6, Application US/08659251			
Patent No. 5883081			
GENERAL INFORMATION:			
APPLICANT: Kraus, Guenter			
APPLICANT: Wong-Staal, Flossie			
APPLICANT: Talbot, Randy			

CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA: US/07/941,372  
CC APPLICATION NUMBER: US/07/941,372  
CC FILING DATE: 19920902  
CC CLASSIFICATION: 424  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Bak, Mary E.  
CC REGISTRATION NUMBER: 31,215  
CC REFERENCE/DOCKET NUMBER: INDUS1  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (215) 540-9206  
CC TELEFAX: (215) 540-5818  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 199 amino acids  
CC TYPE: AMINO ACID  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 199 AA; 21429 MW; 188641 CN;  
  
Query Match 30.6%; Score 60; DB 1; Length 199;  
Best Local Similarity 70.0%; Pred. No. 9.90e+01;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
Db 129 LQARLDRLLR 138  
QY 7 LQGRLLRLQ 16  
||:||||:  
||:||||:  
  
RESULT 14  
ID US-08-792-019B-8 STANDARD; PRT; 199 AA.  
XX AC xxxxxx  
XX AC  
XX DT  
XX DE  
XX SEQUENCE 8, Application US/08792019B  
XX Patent No. 5741772  
XX GENERAL INFORMATION:  
XX APPLICANT: CHANG, MING-SHI  
XX TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
XX NUMBER OF SEQUENCES: 16  
XX CORRESPONDENCE ADDRESS:  
XX ADDRESSEE: AMGEN INC.  
XX STREET: 1840 DEHAVILLAND DRIVE  
XX CITY: THOUSAND OAKS  
XX STATE: CA  
XX COUNTRY: USA  
XX ZIP: 91320  
XX COMPUTER READABLE FORM:  
XX MEDIUM TYPE: Floppy disk  
XX OPERATING SYSTEM: IBM PC compatible  
XX SOFTWARE: PatentIn Release #1.0, Version #1.30  
XX CURRENT APPLICATION DATA:  
XX APPLICATION NUMBER: US/08/792,019B  
XX FILING DATE: 03-FEB-1997  
XX CLASSIFICATION: 514  
XX ATTORNEY/AGENT INFORMATION:  
XX NAME: COOK, ROBERT R.  
XX REGISTRATION NUMBER: 31,602  
XX REFERENCE/DOCKET NUMBER: A-442  
XX INFORMATION FOR SEQ ID NO: 8:  
XX SEQUENCE CHARACTERISTICS:  
XX LENGTH: 199 amino acids  
XX TYPE: amino acid  
XX STRANDEDNESS: single  
XX TOPOLOGY: linear  
XX MOLECULE TYPE: protein

CC FEATURE:  
CC NAME/KEY: Protein  
CC LOCATION: 1..178  
CC FEATURE:  
CC NAME/KEY: Region  
CC LOCATION: -21..0  
CC SEQUENCE 199 AA; 21429 MW; 188641 CN;  
  
Query Match 30.6%; Score 60; DB 1; Length 199;  
Best Local Similarity 70.0%; Pred. No. 9.90e+01;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
Db 129 LQARLDRLLR 138  
QY 7 LQGRLLRLQ 16  
||:||||:  
||:||||:  
  
RESULT 15  
ID US-07-949-516A-4 STANDARD; PRT; 199 AA.  
XX AC xxxxxx  
XX AC  
XX DT  
XX DE  
XX SEQUENCE 4, Application US/07949516A  
XX Patent No. 5700664  
XX GENERAL INFORMATION:  
XX APPLICANT: Yang, Yu-Chung  
XX APPLICANT: Bennett, Frances  
XX TITLE OF INVENTION: A MAMMALIAN CYTOKINE, IL-11  
XX NUMBER OF SEQUENCES: 4  
XX CORRESPONDENCE ADDRESS:  
XX ADDRESSEE: LEGAL AFFAIRS  
XX STREET: 87 Cambridgepark Drive  
XX CITY: Cambridge  
XX STATE: MA  
XX COUNTRY: USA  
XX ZIP: 02140  
XX COMPUTER READABLE FORM:  
XX MEDIUM TYPE: Floppy disk  
XX OPERATING SYSTEM: IBM PC compatible  
XX SOFTWARE: PatentIn Release #1.0, Version #1.25  
XX CURRENT APPLICATION DATA:  
XX APPLICATION NUMBER: US/07/949,516A  
XX FILING DATE: 19-NOV-1992  
XX CLASSIFICATION: 530  
XX ATTORNEY/AGENT INFORMATION:  
XX NAME: Meinert, M. C.  
XX REGISTRATION NUMBER: 31,544  
XX REFERENCE/DOCKET NUMBER: 5174BPCT  
XX INFORMATION FOR SEQ ID NO: 4:  
XX SEQUENCE CHARACTERISTICS:  
XX LENGTH: 199 amino acids  
XX TYPE: amino acid  
XX TOPOLOGY: linear  
XX MOLECULE TYPE: protein  
XX SEQUENCE 199 AA; 21429 MW; 188641 CN;  
  
Query Match 30.6%; Score 60; DB 1; Length 199;  
Best Local Similarity 70.0%; Pred. No. 9.90e+01;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
Db 129 LQARLDRLLR 138  
QY 7 LQGRLLRLQ 16  
||:||||:  
||:||||:  
  
Search completed: Fri Aug 20 20:59:08 1999  
Job time : 8 secs.

\*\*\*\*\*

W P E R L A (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 20:56:23 1999; MasPar time 4.60 Seconds  
129.426 Million cell updates/sec

Tabular output not generated.

Title: >US-08-938-548B-4  
Description: (1-28) from US08938548B.pep  
Perfect Score: 196  
Sequence: 1 RSGPPGQLRQLRLLQASGNHAAIGLTM 28  
Scoring table: PAM 150  
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 22.064; Variance 88.246; scale 0.250

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description	Pred. No.
1	196	100.0	131	33	W61381 Human HFGAN72 recepto	7.83e-11
2	190	96.9	123	33	W61383 Mouse HFGAN72 recepto	3.06e-10
3	190	96.9	130	30	W50158 Mouse hypocretin 35.	3.06e-10
4	190	96.9	130	30	W50157 Rat hypocretin 35.	3.06e-10
5	190	96.9	130	33	W61382 Rat HFGAN72 receptor	3.06e-10
6	68	34.7	716	18	R99737 Retinoid X receptor i	6.30e+01
7	67	34.2	1477	13	R76761 S. cerevisiae scaur2R	7.60e+01
8	67	34.2	1477	22	W10424 Saccharomyces cerevis	7.60e+01
9	67	34.2	1841	26	W22605 Tylosin synthase OR	7.60e+01
10	67	34.2	2192	23	W21732 LexA/NUMA fusion prot	7.60e+01
11	67	34.2	2272	23	W21731 GAL4/HA/NUMA fusion p	7.60e+01
12	67	34.2	4630	23	W19629 Streptomyces venezuel	7.60e+01
13	66	33.7	10	4	R02135 "p33" N-terminal (2).	9.16e+01
14	65	33.2	381	38	W21135 Human 3-hydroxyisobut	1.10e+02
15	63	32.1	32	1	R00579 New polypeptide based	1.60e+02
16	63	32.1	248	2	R05091 Vector PSP 35K-1A-10	1.60e+02

17	63	32.1	248	1	R06331 Human alveolar surfac	1.60e+02
18	63	32.1	248	3	P60442 Plasmid PASPCq-SV(10)	1.60e+02
19	63	32.1	248	1	R04215 Human 32K ASP encoded	1.60e+02
20	63	32.1	248	1	P82980 Sequence deduced from	1.60e+02
21	63	32.1	248	2	P70663 35kd pulmonary surfac	1.60e+02
22	63	32.1	248	1	R04216 Human 32K ASP encoded	1.60e+02
23	63	32.1	248	1	P80694 Sequence deduced from	1.60e+02
24	63	32.1	248	2	P70662 35kd pulmonary surfac	1.60e+02
25	63	32.1	248	3	P60666 Genomic sequence of h	1.60e+02
26	63	32.1	248	3	P60665 Sequence of human alv	1.60e+02
27	63	32.1	248	3	P60441 Plasmid PASPC-SV(10)	1.60e+02
28	63	32.1	271	3	P60661 Genomic sequence of h	1.60e+02
29	63	32.1	271	1	R04212 Human 32K alveolar su	1.60e+02
30	63	32.1	271	1	R04217 Human 32K ASP encoded	1.60e+02
31	63	32.1	550	29	W45513 Ecdysone receptor lig	1.60e+02
32	63	32.1	550	29	W39139 Drosophila ecdysone r	1.60e+02
33	63	32.1	746	28	W33654 Modified ecdysone rec	1.60e+02
34	63	32.1	746	28	W33655 Modified ecdysone rec	1.60e+02
35	63	32.1	878	6	R32889 DHR23alpha protein.	1.60e+02
36	63	32.1	878	3	R13793 Ecdysone receptor.	1.60e+02
37	63	32.1	1041	28	W33656 Modified ecdysone rec	1.60e+02
38	62	31.6	176	23	W13056 HIV-2 provirus-encode	1.92e+02
39	62	31.6	510	9	R49835 Thermus aquaticus hea	1.92e+02
40	62	31.6	2509	24	W32881 Protein (OA-519) cros	1.92e+02
41	60	30.6	178	18	W02202 Human interleukin-11	2.75e+02
42	60	30.6	495	30	W46904 A human mutant alanin	2.75e+02
43	60	30.6	530	14	R76480 Virulence-associated	2.75e+02
44	60	30.6	551	38	W83404 Human KM-102-derived	2.75e+02
45	60	30.6	3724	26	W23718 Platenolide synthase	2.75e+02

ALIGNMENTS

RESULT 1  
ID W61381 standard; Protein; 131 AA.

AC W61381; (first entry)  
DT 02-OCT-1998  
DE Human HFGAN72 receptor protein.  
KW HFGAN72 receptor; eating disorders; renal disease; heart failure;  
KW sexual dysfunction; HIV; cancer; Parkinson's disease; urinary retention;  
KW osteoporosis; angina pectoris; myocardial infarction; psychotic disorder;  
KW neurological disorder.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Region 34..66  
FT Region /note= "Ligand 72A"  
FT Region /note= "Ligand 72B"

EP-849361-A2.  
24-JUN-1998.  
17-DEC-1997; 310216  
26-SEP-1997; US-939093.  
17-DEC-1996; US-033604.  
19-MAR-1997; US-820519.  
02-JUL-1997; US-887382.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PI Bergsma DJ, Brooks DP, Gellal M, Wilson S, Yanagisawa M;  
DR WPI; 98-324672/29.  
DR N-PSDB; V28138.  
PT HFGAN72 receptor ligands - and corresponding DNA, agonists,  
antibodies, antagonists, etc.  
PS Claim 2; Fig 2; 35pp; English.  
CC The HFGAN72 receptor protein contains two ligands whose antagonists can  
be used for treating obesity, diabetes, anorexia nervosa, bulimia,  
cachexia, chronic renal failure, renal dysfunction, congestive heart failure,  
impaired glucose tolerance and sexual dysfunction. The agonist is  
useful for treating anorexia nervosa, bulimia and cachexia. The HFGAN72  
receptor ligand is useful for treating e.g. bacterial, fungal, protozoan  
and viral infections, particularly infections caused by HIV-1 or HIV-2,  
pain, cancers, anorexia nervosa, bulimia, cachexia, obesity, diabetes,  
asthma, Parkinson's disease, both acute and congestive heart failure,  
hypotension, hypertension, urinary retention, osteoporosis, angina

CC antagonists, in therapy, to detect Ab and to isolate cognate  
 CC receptors. Oligonucleotides based on H35 cDNA can be used to detect  
 CC the hypocretin gene or its RNA transcript, and as antisense agents  
 CC for inhibiting gene expression. H35 cDNA can also be used for  
 CC recombinant protein production. The Ab can be used to detect or  
 CC quantify hypocretin proteins and as a therapeutic inhibitor.  
 CC Sequence 130 AA;

Query Match 96.9%; Score 190; DB 30; Length 130;  
 Best Local Similarity 92.9%; Pred. No. 3.06e-10;  
 Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 69 rpgppglqgrllqangnhaagiltm 96  
 QY 1 RSGPPGLQGRLLQASGNHAAIGILTM 28

RESULT 5  
 ID W61382 standard; Protein; 130 AA.  
 AC W61382;  
 DT 02-OCT-1998 (first entry)  
 DE Rat HFGAN72 receptor protein.  
 KW HFGAN72 receptor; eating disorders; renal disease; heart failure;  
 KW sexual dysfunction; HIV; cancer; Parkinson's disease; urinary retention;  
 KW osteoporosis; angina pectoris; myocardial infarction; psychotic disorder;  
 KW neurological disorder.  
 OS Rattus sp.  
 FH Key Location/Qualifiers  
 FT Region 33..65  
 FT /note= "Ligand 72A"  
 FT /note= "Ligand 72B"

FT EP-849361-A2.  
 PN 24-JUN-1998.  
 PD 17-DEC-1997; 310216.  
 PF 26-SEP-1997; US-939093.  
 PR 17-DEC-1996; US-033604.  
 PR 19-MAR-1997; US-820519.  
 PR 02-JUL-1997; US-887382.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Bergsma DJ, Brooks DP, Gellai M, Wilson S, Yanagisawa M;  
 DR WPI: 98-324672/29.  
 DR N-PSDB; V28139.  
 DR HFGAN72 receptor ligands - and corresponding DNA, agonists,  
 PT antibodies, antagonists, etc.  
 PS Claim 5; Fig 4; 35pp; English.

CC The HFGAN72 receptor protein contains two ligands whose antagonists can be  
 CC used for treating obesity, diabetes, anorexia nervosa, bulimia, cachexia,  
 CC chronic renal failure, renal disease, congestive heart failure, impaired  
 CC glucose tolerance and sexual dysfunction. The agonist is useful for  
 CC treating anorexia nervosa, bulimia and cachexia. The HFGAN72 receptor  
 CC ligand is useful for treating e.g. bacterial, fungal, protozoan and viral  
 CC infections, particularly infections caused by HIV-1 or HIV-2, pain,  
 CC cancers, anorexia nervosa, bulimia, cachexia, obesity, diabetes, asthma,  
 CC Parkinson's disease, both acute and congestive heart failure.  
 CC hypotension, hypertension, urinary retention, osteoporosis, angina  
 CC pectoris, myocardial infarction, ulcers, asthma, allergies, benign  
 CC prostatic hypertrophy, chronic renal failure, renal disease, impaired  
 CC glucose tolerance, sexual dysfunction and psychotic and neurological  
 CC disorders including anxiety, schizophrenia, manic depression, delirium,  
 CC dementia, severe mental retardation and dyskinesias such as Huntington's  
 CC disease or Gilles de la Tourette's syndrome.  
 CC Sequence 130 AA;

Query Match 96.9%; Score 190; DB 33; Length 130;  
 Best Local Similarity 92.9%; Pred. No. 3.06e-10;  
 Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 69 rpgppglqgrllqangnhaagiltm 96  
 QY 1 RSGPPGLQGRLLQASGNHAAIGILTM 28

## RESULT 6

ID R99737 standard; Protein; 716 AA.  
 AC R99737;  
 DT 27-SEP-1996 (first entry)  
 DE Retinoid X receptor interacting protein R1P110.  
 KW Retinoid X receptor interacting protein; RXR; R1P110.  
 OS Mus sp.

PN W09621677-A1.  
 PD 18-JUL-1996.  
 PF 08-DEC-1995; U16311.  
 PR 13-JAN-1995; US-372652.  
 PA (GEHO ) GEN HOSPITAL CORP.  
 PI Choi H, Moore D, Seol W;  
 DR WPI: 96-342241/34.  
 DR N-PSDB; T31030.  
 PT Retinoid X receptor (RXR) interacting protein (RIP) - useful to  
 PT modulate or mediate RXR function, anti-RIP antibodies can be used to  
 PT determine RIP subcellular distribution patterns  
 PS Claim 2; Page 48-50; 90pp; English.  
 CC Mouse retinoid X receptor (RXR) interacting protein R1P110 (R99737)  
 CC is a candidate transcriptional co-activator. It was identified using  
 CC an in vivo interaction trap system for the isolation of proteins that  
 CC physically interact with RXRs, esp. with the ligand binding domain of  
 CC human RXR alpha. Recombinant R1P110 can be obtd. using a cDNA clone  
 CC (T31930) obtd. from a mouse liver library. R1Ps (see also R99735-36  
 CC and R99738-39) can be used to modulate or mediate RXR function, and  
 CC may be used therapeutically or to raise antibodies.  
 CC Sequence 716 AA;

Query Match 34.7%; Score 68; DB 18; Length 716;  
 Best Local Similarity 44.4%; Pred. No. 6.30e+01;  
 Matches 8; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Db 621 qgrllsllqseqshrttqv 638  
 QY 8 QGRLLQLLQASGNHAAIGI 25

## RESULT 7

ID R67691 standard; Protein; 1477 AA.  
 AC R67691;  
 DT 04-AUG-1995 (first entry)  
 DE S. cerevisiae scaur2R gene product.  
 KW Aureobasidin-resistance; aureobasidin-sensitivity; scaur2R gene;  
 KW mycosis; antimycotic; fungicide; diagnosis; therapy.  
 OS Saccharomyces cerevisiae DKD5D.

PN AU9463129-A.  
 PD 01-DEC-1994.  
 PF 16-MAY-1994; 063129.  
 PR 24-MAY-1993; JP-142523.  
 PR 28-DEC-1993; JP-348893.  
 PA (TAKI ) TAKARA SHUZO CO LTD.  
 PI Kato I, Okada T, Takesako K;  
 DR WPI: 95-023286/04.  
 DR N-PSDB; Q75956.

PT New genes regulating aureobasidin sensitivity - used to develop  
 PT prods. for the diagnosis and treatment of diseases such as  
 PT mycoses

PS Disclosure; Page 77-84; 110pp; English.  
 CC Saccharomyces cerevisiae DKD5D, which is sensitive to aureobasidin,  
 CC was mutagenized with EMS and genomic libraries of 2 resistant  
 CC strains were prepared. Resistant gene scaur2R (Q75954) was isolated  
 CC from 1 library, and resistant gene scaur2R (Q75956) from the other.  
 CC A DNA fragment of scaur1 was used as a probe to isolate a  
 CC sensitive gene, scaur1S (Q75955), from the sensitive strain. A  
 CC T852A mutation leading to a F158Y conversion gave rise to the  
 CC resistance of scaur1R.  
 CC Sequence 1477 AA;

Query Match 34.2%; Score 67; DB 13; Length 1477;  
 Best Local Similarity 40.0%; Pred. No. 7.60e+01;  
 Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;



Db 838 ragrkglearlqlgea 854  
 I:|:|:|:|:|:|:  
 QY 1 RSGPPGLQRLQLLQA 17

## RESULT 11

ID W21731 standard; Protein; 2272 AA.  
 AC W21731;  
 DT 01-OCT-1997 (first entry)  
 DE GAL4/HA/NUMA fusion protein.  
 KW NIP-1; NIP-2; NUMA; nuclear mitotic apparatus; NUMA interacting protein;  
 KW cell division; proliferation; antibody; Ab; detection;  
 KW malignant cell growth.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT domain 1..147  
 FT FT /label= GAL4\_DNA\_binding\_domain  
 FT peptide 148..174  
 FT FT /label= Hemagglutinin-epitope  
 FT protein 175..2272  
 FT FT /label= Residues\_18-2116\_of\_NUMA  
 FT region 365..1864  
 FT FT /label= Coiled\_coil\_region  
 PN WO9640917-A1.  
 PD 19-DEC-1996.  
 PF 07-JUN-1996; U09504.  
 PR 07-JUN-1995; US-478408.  
 PA (UYXA ) UNIV YALE.  
 PI McPherson SMG, Snyder MP;  
 DR WPI: 97-077270/07.  
 DR N-PSDB: T77782.  
 PT New nucleic acid encoding nuclear mitotic appts. interacting  
 PT proteins - useful for modulating cell division and proliferation and  
 PT in diagnosis  
 PS Claim 14; Page 28-36; 78pp; English.  
 CC The sequences given in W21731-32 represent fusion proteins which contain  
 CC NUMA (nuclear mitotic apparatus). The fusion proteins were used in  
 CC the identification of NUMA interacting proteins (NIP's) (see also  
 CC W21729-30). Compounds which interfere with the interaction of NUMA  
 CC with a known NIP are used to modulate cell division and/or proliferation.  
 CC Ab, raised conventionally using NIP-1 or -2 as immunogen, are used to  
 CC detect NIP (or their complexes) and to block their activity for  
 CC diagnostic or therapeutic use, e.g. to detect defective NUMA or NIP  
 CC which may be markers for aberrant (including malignant) cell growth  
 CC (which can also be detected by nucleic acid sequencing). Also where  
 CC malignancy is related to defects in NUMA or NIP, it can be treated by  
 CC administration of the appropriate functional protein.  
 SQ Sequence 2272 AA;

Query Match 34.2%; Score 67; DB 23; Length 2272;  
 Best Local Similarity 52.9%; Pred. No. 7.60e+01;  
 Matches 9; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 918 ragrkglearlqlgea 934  
 I:|:|:|:|:|:|:  
 QY 1 RSGPPGLQRLQLLQA 17

## RESULT 12

ID W19629 standard; Protein; 4630 AA.  
 AC W19629;  
 DT 01-SEP-1997 (first entry)  
 DE Streptomyces venezuelae polyketide synthase.  
 KW Polyketide synthase; polyhydroxyalkanoate monomer synthase;  
 KW polyhydroxybutyrate; biodegradable polymer; vep gene;  
 KW metabolic engineering.  
 OS Streptomyces venezuelae.  
 PN WO9722711-A1.  
 PD 26-JUN-1997.  
 PF 18-DEC-1996; U20119.  
 PR 19-DEC-1995; US-008847.  
 PA (MINU ) UNIV MINNESOTA.  
 PI Sherman DH, Williams MD, Xue Y;

DR WPI: 97-341701/31.  
 DR N-PSDB: T68715.  
 PT Expression cassettes for production of polyhydroxyalkanoate(s) -  
 PT provide wide range of biodegradable polymers for medical or  
 PT industrial use  
 PS Claim 55; Fig 23; 91pp; English.  
 CC 3 Polypeptide sequences (W19629-30 and W00918) can be deduced from  
 CC the vep ORF1 polyketide synthase (PKS) gene cluster (T68715) of  
 CC Streptomyces venezuelae. The sequence data indicate that the PKS  
 CC gene cluster encodes a polyene of 12 carbons. The vep gene cluster  
 CC contains 5 PKS modules, plus a 5' loading module and a 3' end  
 CC domain. Each of the sequenced modules includes a keto-ACP, an  
 CC acyltransferase, a dehydratase, a keto-reductase and an acyl carrier  
 CC protein domain. A novel expression cassette encoding the first  
 CC module from the vep gene cluster and module 7 from the Streptomyces  
 CC tyLP gene cluster has polyhydroxyalkanoate (PHA) monomer synthase  
 CC activity and can be used for PHA prodn. in host (esp. insect) cells  
 CC for use as a biodegradable polymer.  
 SQ Sequence 4630 AA;

Query Match 34.2%; Score 67; DB 23; Length 4630;  
 Best Local Similarity 33.3%; Pred. No. 7.60e+01;  
 Matches 8; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

Db 4302 rdtapaiaahlaellatardhpgg 4325  
 I:|:|:|:|:|:|:  
 QY 1 RSGPPGLQRLQLLQASGNHAAG 24

## RESULT 13

ID R20235 standard; Protein; 10 AA.  
 AC R20235;  
 DT 24-APR-1992 (first entry)  
 DE "p33" N-terminal (2).  
 KW MAP; lymphocyte; IL-2; CTL; polymorphism.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_difference 6 /note= "polymorphism - see CC"  
 FT FT  
 PN WO9200329-A.  
 PD 09-JAN-1992.  
 PF 27-JUN-1991; U04588.  
 PR 27-JUN-1990; US-544862.  
 PA (BIOG-) BIOGEN INC.  
 PA (REGC ) UNIV OF CALIFORNIA.  
 PI Browning J, Ware CF;  
 DR WPI: 92-041521/05.

PT New membrane associated protein, p33 - forms complex with  
 PT lymphotoxin, useful as antiinflammatory agent, tumour growth  
 PT inhibitor, T-cell inhibitor or activator  
 PS Claim 1; Page 60; 75pp; English.  
 CC A protein, designated p33, is found on the surface of several types  
 CC of lymphocyte cells, including OKT3-stimulated primary T-cells,  
 CC antigen-specific IL-2 dependent CTL clones, and a PMA-stimulated  
 CC human T-cell hybridoma, II-23.D7. It forms a novel complex with  
 CC lymphotoxin (LT). The N-terminal of the p33 protein contains the  
 CC sequence represented in R20234 or R20235. The sequence at the 6th  
 CC cycle appeared to be a mixt. of both G and L indicating possible  
 CC polymorphism. The p33 protein may have one of these sequences or  
 CC both.  
 SQ Sequence 10 AA;

Query Match 33.7%; Score 66; DB 4; Length 10;  
 Best Local Similarity 88.9%; Pred. NO. 9.16e+01;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 glegrlql 9  
 I:|:|:|:|:|:|:  
 QY 6 GLOGRLQL 14

## RESULT 14

ID W81135 standard; Protein; 381 AA.

\*\*\*\*\*

WATERBURY

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:00:11 1999; MasPar time 7.71 Seconds  
358.518 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-938-548B-6

Description: (1-130) from US08938548B.pep

Perfect Score: 954  
Sequence: 1 MNLPTKVPWAAVTLILL.....GRRCPATATAPRGGSRV 130

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 30.024; Variance 131.695; scale 0.228

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	954	100.0	130	30	W50157	5.23e-79
2	954	100.0	130	33	W61382	5.23e-79
3	891	93.4	130	30	W50158	6.08e-73
4	853	89.0	123	33	W61383	2.73e-69
5	782	82.0	131	33	W61381	1.76e-62
6	113	11.8	378	39	W84185	4.63e-01
7	113	11.8	400	39	W84186	4.63e-01
8	113	11.8	400	39	W84180	4.63e-01
9	113	11.8	400	32	W65116	4.63e-01
10	113	11.8	400	32	W65117	4.63e-01
11	113	11.8	400	28	W37463	4.63e-01
12	96	10.1	556	28	W27598	9.01e-00
13	96	10.1	556	2	R11448	9.01e-00
14	96	10.1	601	2	R11449	9.01e-00
15	96	10.1	601	28	W27599	9.01e-00
16	96	10.1	683	2	R11150	9.01e-00

17	96	10.1	683	28	W27600	Human fibulin type 1	9.01e+00
18	96	10.1	703	28	W27601	Human fibulin type 1	9.01e+00
19	94	9.9	17	30	W50160	Rat hypocretin 35 pep	1.27e+01
20	91	9.5	440	25	W24789	Human lecithin-choles	2.10e+01
21	91	9.5	440	2	F70134	Natural recombinant h	2.10e+01
22	90	9.4	566	30	W47029	Human N-proteinase (7	2.48e+01
23	90	9.4	1211	30	W47028	Human N-proteinase (1	2.48e+01
24	89	9.3	58	37	W75074	Human secreted protei	2.93e+01
25	89	9.3	190	10	R50012	Truncated Platelet de	2.93e+01
26	89	9.3	216	12	R68617	Human PDGF-B precurs	2.93e+01
27	89	9.3	220	8	R40965	rPDGF B.	2.93e+01
28	89	9.3	241	10	R50002	Platelet-derived Grow	2.93e+01
29	89	9.3	241	12	R63459	Recombinant platelet	2.93e+01
30	89	9.3	241	10	R50009	Platelet-derived Grow	2.93e+01
31	89	9.3	241	8	R40964	PDGF Bc-sis.	2.93e+01
32	89	9.3	241	1	P80596	Recombinant platelet	2.93e+01
33	89	9.3	241	33	W64065	Chimeric rhpDGF-B pro	2.93e+01
34	88	9.2	226	37	W72902	Mycobacterium tubercu	3.46e+01
35	88	9.2	252	4	R21708	HSV-1 (CVG-2) ICP34.5	3.46e+01
36	88	9.2	254	12	R64190	Human 4-1BB-L polypep	3.46e+01
37	88	9.2	254	25	W26657	Human 4-1BB ligand.	3.46e+01
38	87	9.1	111	7	R35515	Tryptophan aporepress	4.08e+01
39	87	9.1	201	4	R21443	Sequence encoded by p	4.08e+01
40	87	9.1	226	2	R22673	v-sis protein p28sis.	4.08e+01
41	87	9.1	241	12	R63472	Recombinant platelet	4.08e+01
42	87	9.1	241	1	P80597	Cv-sis gene encoded p	4.08e+01
43	87	9.1	271	8	R40963	PDGF Bv-sis.	4.08e+01
44	87	9.1	271	1	P80595	Recombinant platelet	4.08e+01
45	87	9.1	438	31	W57042	Human aspartic protea	4.08e+01

#### ALIGNMENTS

RESULT 1  
ID W50157 standard; Protein; 130 AA.  
AC W50157;  
DT 09-JUL-1998 (first entry)  
DE Rat hypocretin 35.  
KW Rat; hypocretin 35; H35; treatment; neurological disease;  
KW homeostatic dysfunction;  
KW homeostatic regulatory hormone production.  
OS Rattus rattus.  
PN WO9805352-A1.  
PD 12-FEB-1998.  
PF 01-AUG-1997; UI3657.  
PR 02-AUG-1996; US-023220.  
PA (SCRI ) SCRIPPS RES INST.  
PI Bloom FE, Danielson PE, De Lecea L, Foye PE, Gautvik KM,  
PI Gautvik VT, Kilduff TS, Sutcliffe JG;  
DR WPI: 98-145352/13.  
DR N-PSDB; V18865.  
PT Nucleic acid encoding hypocretin of rat and mouse - useful for  
PT diagnosis and treatment of neurological disease, homeostatic  
PT dysfunction etc., also sequence for calmodulin kinase-like protein  
PS Claim 1; Page 81; 111pp; English.  
CC The present sequence is rat hypocretin 35 (H35), which is  
CC involved in lowering body temperature and reducing food intake.  
CC Modulation of the H35 receptor can be used in the treatment of  
CC neurological disease or homeostatic dysfunction, or to control  
CC homeostatic regulatory hormone production. Hypocretin proteins can  
CC be used to raise antibodies (Ab), to identify specific agonists or  
CC antagonists, in therapy, to detect Ab and to isolate cognate  
CC receptors. Oligonucleotides based on H35 cDNA can be used to detect  
CC the hypocretin gene or its RNA transcript, and as antisense agents  
CC for inhibiting gene expression. H35 cDNA can also be used for  
CC recombinant protein production. The Ab can be used to detect or  
CC quantify hypocretin proteins and as a therapeutic inhibitor.  
SQ Sequence 130 AA;

Query Match 100.0%; Score 954; DB 30; Length 130;  
Best Local Similarity 100.0%; Pred. No. 5.23e-79;  
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

PN EP-849361-A2.
PD 24-JUN-1998.
PF 17-DEC-1997; 310216.
PR 26-SEP-1997; US-939093.
PR 17-DEC-1996; US-033604.
PR 19-MAR-1997; US-820519.
PR 02-JUL-1997; US-887382.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PI Bergsma DJ, Brooks DP, Gellai M, Wilson S, Yanagisawa M;
DR HFGAN72 receptor ligands - and corresponding DNA, agonists,
PT antibodies, antagonists, etc.
PS Claim 5; Fig 2; 35pp; English.
CC The HFGAN72 receptor protein contains two ligands whose antagonists can
CC be used for treating obesity, diabetes, anorexia nervosa, bulimia, cachexia,
CC chronic renal failure, renal disease, congestive heart failure, impaired
CC glucose tolerance and sexual dysfunction. The agonist is useful for
CC treating anorexia nervosa, bulimia and cachexia. The HFGAN72 receptor
CC ligand is useful for treating e.g. bacterial, fungal, protozoan and viral
CC infections, particularly infections caused by HIV-1 or HIV-2, pain,
CC cancers, anorexia nervosa, bulimia, cachexia, obesity, diabetes, asthma,
CC Parkinson's disease, both acute and congestive heart failure,
CC hypotension, myocardial infarction, ulcers, asthma, allergies, benign
CC pectoris, chronic renal failure, renal disease, osteoporosis, angina
CC prostate hypertrophy, chronic renal failure, renal disease, impaired
CC glucose tolerance, sexual dysfunction and psychotic and neurological
CC disorders including anxiety, schizophrenia, manic depression, delirium,
CC dementia, severe mental retardation and dyskinesias such as Huntington's
CC disease or Gilles de la Tourette's syndrome.
CC Sequence 133 AA;
SQ

Query Match 89.4%; Score 853; DB 33; Length 123;
Best Local Similarity 95.1%; Pred. No. 2.73e-69;
Matches 117; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 1 vpwaaavtlllllppallslgvdqplpdcrcqtkscrllyellhgaghnhaagiltlgk 60
QY 8 VFWAAVTLTLLLLPPALLSLGVDAQPLPDCRCQTKSCRLYELLHGAGNHAAGILT LGK 67

Db 61 rrpqpgqlgrllqanqhaagiltmrragaelphpcsggrgcvtttallaprgg 120
QY 68 RPPGPGQLGRLLQANQHAAGILTMRRAGAELEPYPCGRRCTATATAPRGG 127

Db 121 sgV 123
QY 128 SRV 130

RESULT 5
ID W61381 standard; Protein; 131 AA.
AC W61381;
DT 02-OCT-1998 (first entry)
DE Human HFGAN72 receptor protein.
KW HFGAN72 receptor; eating disorders; renal disease; heart failure;
KW sexual dysfunction; HIV; cancer; Parkinson's disease; urinary retention;
KW osteoporosis; angina pectoris; myocardial infarction; psychotic disorder;
KW neurological disorder.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Region 34..66
FT /note= "Ligand 72A"
FT Region 70..97
FT /note= "Ligand 72B"
FT
PN EP-849361-A2.
PD 24-JUN-1998.
PF 17-DEC-1997; 310216.
PR 26-SEP-1997; US-939093.
PR 17-DEC-1996; US-033604.
PR 19-MAR-1997; US-820519.
PR 02-JUL-1997; US-887382.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PI Bergsma DJ, Brooks DP, Gellai M, Wilson S, Yanagisawa M;
DR HFGAN72 receptor ligands - and corresponding DNA, agonists,
PT antibodies, antagonists, etc.
PS Claim 5; Fig 2; 35pp; English.
CC The HFGAN72 receptor protein contains two ligands whose antagonists can
CC be used for treating obesity, diabetes, anorexia nervosa, bulimia, cachexia,
CC chronic renal failure, renal disease, congestive heart failure, impaired
CC glucose tolerance and sexual dysfunction. The agonist is useful for
CC treating anorexia nervosa, bulimia and cachexia. The HFGAN72 receptor
CC ligand is useful for treating e.g. bacterial, fungal, protozoan and viral
CC infections, particularly infections caused by HIV-1 or HIV-2, pain,
CC cancers, anorexia nervosa, bulimia, cachexia, obesity, diabetes, asthma,
CC Parkinson's disease, both acute and congestive heart failure,
CC hypotension, myocardial infarction, ulcers, asthma, allergies, benign
CC pectoris, chronic renal failure, renal disease, osteoporosis, angina
CC prostate hypertrophy, chronic renal failure, renal disease, impaired
CC glucose tolerance, sexual dysfunction and psychotic and neurological
CC disorders including anxiety, schizophrenia, manic depression, delirium,
CC dementia, severe mental retardation and dyskinesias such as Huntington's
CC disease or Gilles de la Tourette's syndrome.
CC Sequence 133 AA;
SQ

Query Match 82.0%; Score 782; DB 33; Length 131;
Best Local Similarity 85.2%; Pred. No. 1.76e-62;
Matches 109; Conservative 10; Mismatches 7; Indels 2; Gaps 2;

Db 1 mnlpstkvswaavtlllllppallslgaaqplpdcrcqtkscrllyellhgaghnha 60
QY 1 MNLPSKVPWAAVTLTLLLL-PPALLSLGVDAQPLPDCRCQTKSCRLYELLHGAGNHA 59

Db 61 agiltlgrsppqlgrllqanqhaagiltmrragaelphpcsggrgcvtttallaprgg 120
QY 60 AGILTLGRSPPQLGRLLQANQHAAGILTMRRAGAELEPYPCGRRCTATATA 119

Db 121 asvap-gg 127
QY 120 TALAPRGG 127

RESULT 6
ID W84185 standard; Protein; 378 AA.
AC W84185;
DT 25-MAR-1999 (first entry)
DE Glial cell line-derived neurotrophic factor receptor gamma 1.
KW Glial cell line-derived neurotrophic factor receptor gamma 1;
KW GDNFR-alpha; glial cell line-derived neurotrophic factor; GDNF;
KW neurodegenerative disease; amyotrophic lateral sclerosis; GDNFR-gamma1;
KW Parkinson's disease; schizophrenia; insomnia; tardive dyskinesia;
KW hypertension; pituitary adenomas; hyperprolactinemia; thyroid tumour;
KW renal disorder; kidney failure; gut dysfunction; regeneration;
KW cardiomyocyte; epithelium; hepatocyte.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..31
FT /label= signal_peptide
FT Protein 32..378
FT /label= mature_protein
FT Domain 32..360
FT /note= "extracellular domain"
FT Domain 361..378
FT /note= "transmembrane domain"
FT
PN W09853069-A2.
PD 26-NOV-1998.
PF 20-MAY-1998; U10328.
PR 27-JUN-1997; US-884638.
PR 20-MAY-1997; US-047092.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Gentz RL, Hsu T, Ni J, Ruben SM, Young P;
DR WPI; 99-070150/06.
DR N-PSDB; V99333.
PT New isolated glial cell derived neurotrophic factor receptors - used

```



CC single residue. Embodied within repeats 5, 6, 7 and 8 is the  
 CC consensus sequence for Asp and Asn hydroxylation. The 7th repeat  
 CC contains the consensus O-glycosylation sequence, CXCXPC.  
 CC Immediately following each repeat is a pentapeptide with the  
 CC consensus sequence XD(I/V)(D/N)E. Fibulin binds to the cyto-  
 CC plasmic domain of the beta1 subunit of integrin adhesion  
 CC receptors in a cation-dependent, EDTA-reversible manner. It can  
 CC be used to manipulate adhesion of cells to fibronectin, collagen,  
 CC laminin, and possibly also other proteins. Antibodies reactive  
 CC with the protein have important diagnostic and therapeutic uses.  
 CC See also R11147, R11149 and R11150.  
 SQ Sequence 566 AA;

Query Match 10.1%; Score 96; DB 2; Length 566;

Best Local Similarity 57.1%; Pred. No. 9.01e+00;

Matches 16; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

Db 10 vplpllllglllaagvadvllleacc 37

QY 13 VTLLLLLLPPALLSLGVDAQPLPD-CC 39

RESULT 14

ID R11149 standard; Protein; 601 AA.

AC R11149, 1991 (first entry)

DE Fibulin B.

KW Beta-1 integrin; adhesion; receptor; fibronectin.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 1..29

FT protein /label= signal sequence

FT modified\_site 30..601

FT /label= fibulin B

FT modified\_site 98

FT /label= N-linked glycosylation

FT modified\_site 535

FT modified\_site 339

FT /label= N-linked glycosylation

FT /label= type I motif

FT /label= repeat unit 1

FT /label= repeat unit 2

FT /label= Glu/Asp-rich region

FT /label= type II motif

FT /label= repeat unit 1

FT /label= consensus pentapeptide

FT /label= repeat unit 2

FT /label= consensus pentapeptide

FT /label= repeat unit 3

FT /label= consensus pentapeptide

FT /label= repeat unit 4

FT /label= consensus pentapeptide

FT /label= repeat unit 5

FT /label= consensus pentapeptide

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT /label= repeat unit 6

FT region /label= consensus pentapeptide

FT /label= repeat unit 7

FT /label= consensus pentapeptide

FT /label= repeat unit 8

FT /label= consensus pentapeptide

PN W09102755-A.

PD 07-MAR-1991.

PF 17-AUG-1990; U04662.

PR 18-AUG-1989; US-395773.

PA (LJOL-) LA JOLLA CANCER RES.

PA (AMNA-) AMER NAT RED CROSS.

PI Ruoslahti EI, Argraves WS;

DR WPI: 91-087250/12.

DR N-PSDB: Q11009.

PT Purified fibulin, DNA encoding it and antibodies reactive with it

PT - useful as diagnostic and therapeutic component.

PS Claim 10; Fig 4; 56pp; English.

CC The fibulin A, B and C forms are identical from their N-terminals

CC to a divergence point at posn. 566 (terminal codon in fibulin A)

CC after which they are distinct, encoding polypeptides of 566, 601

CC and 683 residues resp. All three forms are rich in cysteine (11 %)

CC and analysis wrt no. and spacing of the Cys residues revealed two

CC types of repeat motif (I and II). The type I motif, CC(X)12C-

CC (X)9-10C(X)6CC is repeated twice, separated by an imperfect form of

CC the repeat lacking two Cys residues. The same motif is found in

CC complement components C3a, C4a and C5a; the inverse is found in

CC albumin, vitamin D-binding protein and alpha-fetoprotein. The

CC disulphide-stabilized loop structure is thought to be conserved.

CC The type II motif, related to repeats found in epidermal growth

CC factor precursor is a 6-Cys pattern repeated nine times, although

CC the ninth repeat in the A form is incomplete. Four of the repeats,

CC (2, 3, 4 and 9) differ from the EGF-like motif in having a 4-6

CC residue insertion between cysteines 4 and 5 instead of the usual

CC single residue. Embodied within repeats 5, 6, 7 and 8 is the

CC consensus sequence for Asp and Asn hydroxylation. The 7th repeat

CC contains the consensus O-glycosylation sequence, CXCXPC.

CC Immediately following each repeat is a pentapeptide with the

CC consensus sequence XD(I/V)(D/N)E. Fibulin binds to the cyto-

CC plasmic domain of the beta1 subunit of integrin adhesion

CC receptors in a cation-dependent, EDTA-reversible manner. It can

CC be used to manipulate adhesion of cells to fibronectin, collagen,

CC laminin, and possibly also other proteins. Antibodies reactive

CC with the protein have important diagnostic and therapeutic uses.

CC See also R11147, R11148 and R11150.

SQ Sequence 601 AA;

Query Match 10.1%; Score 96; DB 2; Length 601;

Best Local Similarity 57.1%; Pred. No. 9.01e+00;

Matches 16; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

Db 10 vplpllllglllaagvadvllleacc 37

QY 13 VTLLLLLLPPALLSLGVDAQPLPD-CC 39

RESULT 15

ID W27599 standard; Protein; 601 AA.

AC W27599;

DT 12-MAY-1998 (first entry)

DE Human fibulin type 1 isoform (variant B).

KW Human fibulin protein; met-OB protein; hypertension; obesity;

OS isoform; type II diabetes; fibrogenesis.

OS Homo sapiens.

PN W09738014-A1.

PD 16-OCT-1997.

PF 03-APR-1997; U06280.

PR 04-APR-1996; US-627636.

PA (AMGE-) AMGEN INC.

PI Bennett LG;

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

MPsrch_pp  protein - protein database search, using Smith-Waterman algorithm

Run on:      Fri Aug 20 21:02:20 1999;      MasPar time 11.91 seconds
Tabular output not generated.      595.879 Million cell updates/sec

```

Title: >US-08-938-548B-6  
Description: (1-130) from US08938548B.pep  
Perfect Score: 954  
Sequence: 1 MNPSTKVPWAAVTLKLLLL.....

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0% Listing first 45 summaries

Database: sptrembl9

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phage 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 40.229; Variance 88.216; scale 0.456

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	954	100.0	130	11	055232	PREPRO-OREXIN.	1.31e-153
2	894	93.7	130	11	053411	HYPOCRETIN (PREPRO-ORE	3.25e-142
3	782	82.0	131	4	043612	PREPRO-OREXIN.	4.87e-121
4	756	79.2	131	6	077668	PREPRO-OREXIN PRECURS	3.82e-116
5	113	71.8	400	4	060609	GNDF FAMILY RECEPTOR A	5.76e-03
6	106	11.1	679	10	023352	HYPOPHETICAL 74.6 KD P	4.77e-02
7	98	10.3	574	11	035298	ACTOXYACIL HYDROLASE.	4.90e-01
8	98	10.3	641	11	008463	FRIZZLED PROTEIN HOMOL	4.90e-01
9	96	10.1	272	5	077465	LARGE RIBOSOMAL PROTEI	8.63e-01
10	96	10.1	626	11	070421	FRIZZLED-1.	8.63e-01
11	95	10.1	920	2	087342	BETA-(1-3)-GLUCOSYL TR	8.63e-01
12	92	9.6	106	10	041051	PROLINE- AND LEUCINE-R	2.62e+00
13	91	9.5	440	11	035849	LECITHIN-CHOLESTEROL A	3.45e+00
14	90	9.4	833	4	099523	SORTILIN PRECURSOR.	4.53e+00
15	89	9.3	144	4	075779	COLL1A1 AND PDGFB FUSIO	5.93e+00
16	89	9.3	154	4	015186	COLL1A1 AND PDGFB FUSIO	5.93e+00
17	89	9.3	175	4	060895	MRNA ENCODING RAMP2 PR	5.93e+00
18	89	9.3	185	4	015354	C-SIS PROTO-ONCOGENE (	5.93e+00
19	89	9.3	331	13	091640	LEUCINE ZIPPER WITH BAG	5.93e+00
20	89	9.3	1321	4	075129	KIAA0634 PROTEIN (FRAG	5.93e+00

```

RA SCHUELLER C., CHALWATZIS N.;
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RL Arabidopsis thaliana.";
RE NATURE 391:485-488(1998).
DR EMBL; Z97337; E328841; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 679 AA; 74635 MW; B301B713 CRC32;

Query Match      11.1%; Score 106; DB 10; Length 679;
Best Local Similarity 32.3%; Pred No. 4.77e-02;
Matches 20; Conservative 16; Mismatches 22; Indels 4; Gaps 4;

Db    598 KYCRSK-YETIHGNDHNAADVLEAIKREMPAELL-R-ASLRHTNEDQRFLLNVGRSA 654
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY   42 KTCSCRILYELHGAG-NHAAGIILTCKRRPPGLQRLQLRQAANGNHAAGILTMGRRA 100
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db    655 SP 556
QY   101 GA 102

RESULT 7
ID O35298 PRELIMINARY; PRT; 574 AA.
AC O35298;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ACILOXYACYL HYDROLASE.
GN AOAH.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA MUNFORD R.S., FOSMIRE S., VARLEY A.W., STAAB J.F.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF018172; G2529571; -.
DR PFAM; PF00657; Lipase_GDSL; 1.
KW HYDROLASE.
SQ SEQUENCE 574 AA; 65154 MW; B7D6EE69 CRC32;

Query Match      10.3%; Score 98; DB 11; Length 574;
Best Local Similarity 38.1%; Pred. No. 4.90e-01;
Matches 16; Conservative 7; Mismatches 17; Indels 2; Gaps 2;

Db    2 KFPWKVKFTLLLLLSHSLASVPSEDQPDSYSHGQSCLGC 43
      | | | | | | | | | | | | | | | | | | | | |
QY   7 KPWAAY-TLLLLLLPALLSLGVDAQLPLDCCRQKTC-SC 46

RESULT 8
ID Q08463 PRELIMINARY; PRT; 641 AA.
AC Q08463;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE FRIZZLED PROTEIN HOMOLOG 1 (Fz-1).
GN Fz-1.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=OSTEOSARCOMA;
RX MEDLINE; 93094228.
RA CHAN S.D.H., KARPF D.B., FOWLKES M.E., HOOKS M., BRADLEY M.S.,
RA VUONG V., BAMBINO T., LIU M.Y.C., ARNAUD C.D., STREWLER G.J.,
RA NISSENSON R.A..
RT "Two homologs of the Drosophila polarity gene frizzled (fz) are
RT widely expressed in mammalian tissues.";
RL J. BIOL. CHEM. 267:25202-25207(1992).
CC -!- FUNCTION: MAY BE INVOLVED IN TRANSDUCTION AND INTERCELLULAR
```

```

RT purified from human brain by receptor-associated protein affinity chromatography."
RL J. BIOL. CHEM. 272:3599-3605(1997).
DR EMBL; X98248; E246784; -
KW SIGNAL.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 833 SORTILIN.
SQ SEQUENCE 833 AA; 92408 MW; D2E351B9 CRC32;

Query Match 9.4%; Score 90; DB 4; Length 833;
Best Local Similarity 53.3%; Pred. No. 4.53e+00;
Matches 16; Conservative 5; Mismatches 6; Indels 3; Gaps 3;

Db 14 WPHGIGLLILLQLLPSTLSQDLRAPP 43
   ::::::|::|::|::|::|::|::|
QY 10 WA-AVTLILL-LIPALLSLG-VDAQPLP 36

RESULT 15
ID Q75779 PRELIMINARY; PRT; 144 AA.
AC O75779;
DT 01-NOV-1998 (TREMELREL. 08, CREATED)
DT 01-NOV-1998 (TREMELREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMELREL. 08, LAST ANNOTATION UPDATE)
DE COLIAI AND PDGFB FUSION TRANSCRIPT (FRAGMENT).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
CC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA O'BRIEN K.P.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR EMBL; Y15921; E1309800; -
FT NON_TER 1
FT NON_TER 144
SQ SEQUENCE 144 AA; 15626 MW; AF87E707 CRC32;

Query Match 9.3%; Score 89; DB 4; Length 144;
Best Local Similarity 43.6%; Pred. No. 5.93e+00;
Matches 17; Conservative 11; Mismatches 6; Indels 5; Gaps 5;

Db 59 LQRLLHGDFGEEDALNMTSHSGGELESIA-RGRR 96
   |||||::|::|::|::|::|::|::|::|::|::|::|::|
QY 79 LQRLLQAN-GNH-AAGI-LTMGR-RAGAELPYPCPGRR 113

Search completed: Fri Aug 20 21:03:02 1999
Job time : 42 secs.

```



\*\*\*\*\*

W P S R L (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:01:51 1999; MasPar time 6.13 Seconds

Tabular output not generated. 599.747 Million cell updates/sec

Title: >US-08-938-548B-6  
Description: (1-130) from US08938548B.pep

Perfect Score: 954  
Sequence: 1 MNLPSTKVPWAAVTLTLLLL.....GRRCPATATATAPRGGSRV 130

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 41.950; Variance 76.092; scale 0.551

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	103	10.8	567	1	GPV_RAT	1.07e-02
2	101	10.6	131	1	SECR_PIG	2.12e-02
3	96	10.1	438	1	LCAT_MOUSE	1.12e-01
4	96	10.1	566	1	FBLA_HUMAN	1.12e-01
5	96	10.1	601	1	FBLB_HUMAN	1.12e-01
6	96	10.1	683	1	FBLB_HUMAN	1.12e-01
7	96	10.1	703	1	FBLD_HUMAN	1.12e-01
8	93	9.7	1027	1	CAPP_RIFPA	2.97e-01
9	93	9.7	1061	1	ATRIAL_NATRIURETIC_PEP	2.97e-01
10	91	9.5	251	1	C1QB_HUMAN	5.61e-01
11	91	9.5	440	1	LCAT_RAT	5.61e-01
12	91	9.5	440	1	LCAT_HUMAN	5.61e-01
13	89	9.3	241	1	PCGB_HUMAN	1.05e+00
14	89	9.3	331	1	PRP1_HUMAN	1.05e+00
15	89	9.3	602	1	PGH1_RAT	1.05e+00
16	88	9.2	252	1	ICP3_HSV1D	1.43e+00
17	88	9.2	254	1	41BL_HUMAN	1.43e+00
18	88	9.2	497	1	SC14_YARLI	1.43e+00
19	87	9.1	226	1	TS1S_SMSAV	1.94e+00
20	86	9.0	245	1	ICP3_HSVIN	2.64e+00
21	86	9.0	322	1	RLUC_HAEIN	2.64e+00
22	86	9.0	334	1	FEPD_ECOLI	2.64e+00
23	86	9.0	440	1	LCAT_PAPAN	2.64e+00

24	86	9.0	1103	1	CYGD_HUMAN	2.64e+00
25	86	9.0	1663	1	CO3_RAT	2.64e+00
26	85	8.9	236	1	PIC1_BOVIN	3.56e+00
27	85	8.9	241	1	PDGB_MOUSE	3.56e+00
28	85	8.9	255	1	YPE1_RHOU	3.56e+00
29	85	8.9	317	1	LIP1_PSYIM	3.56e+00
30	84	8.8	90	1	VGE_BPHX	4.80e+00
31	84	8.8	90	1	VGE_BFS13	4.80e+00
32	84	8.8	238	1	EFA3_HUMAN	4.80e+00
33	84	8.8	315	1	LIP3_MORSP	4.80e+00
34	84	8.8	319	1	RLUC_ECOLI	4.80e+00
35	84	8.8	384	1	RN_DROME	4.80e+00
36	84	8.8	696	1	LSHR_PIG	4.80e+00
37	84	8.8	1001	1	PTPX_MOUSE	4.80e+00
38	83	8.7	224	1	OXO2_HORVU	6.46e+00
39	83	8.7	235	1	FL3L_HUMAN	6.46e+00
40	83	8.7	281	1	POT1_ECOLI	6.46e+00
41	83	8.7	440	1	LCAT_RABIT	6.46e+00
42	83	8.7	536	1	YABK_ECOLI	6.46e+00
43	83	8.7	1004	1	PTPX_RAT	6.46e+00
44	83	8.7	1109	1	CYGD_CANFA	6.46e+00
45	83	8.7	1310	1	ACE_RABIT	6.46e+00

#### ALIGNMENTS

RESULT 1  
ID GPV\_RAT STANDARD; PRT; 567 AA.  
AC O08770;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE PLATELET GLYCOPROTEIN V PRECURSOR (GPV) (CD42D).  
GN GP5.  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WISTAR; TISSUE=LIVER;  
RX MEDLINE; 97275136.  
RA RAVANAT C., MORALES M., AZORSA D.O., MOOG S., SCHUHLER S.,  
GRUNERT P., LOEW D., VAN DORSELAE A., CAZENAVE J.-P., LANZA F.;  
RT "Gene cloning of rat and mouse platelet glycoprotein V:  
RT identification of megakaryocyte-specific promoters and demonstration  
RT of functional thrombin cleavage."  
RL BLOOD 89:3253-3262(1997).  
CC -!- FUNCTION: THE GPB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND  
CC FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT  
CC PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO  
CC INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A  
CC CRITICAL INITIATING EVENT IN HEMOSTASIS (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN  
CC MANY PROTEINS. NUMBER IN THIS PROTEIN: 15.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z69594; E222201; -.  
CC PFAM; PF00560; LRR; 8.  
DR PLATELET; TRANSMEMBRANE; GLYCOPROTEIN; BLOOD COAGULATION;  
KW REPEAT; LEUCINE-REPEAT; CELL ADHESION; SIGNAL.  
FT SIGNAL 1 16  
FT CHAIN 17 567  
FT DOMAIN 17 522  
FT TRANSMEM 523 543  
FT DOMAIN 544 567  
FT CYTOPLASMIC- (POTENTIAL).

[illegible]

FT	DOMAIN	262	307	EGF-LIKE 3,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	308	355	EGF-LIKE 4,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	356	398	EGF-LIKE 5,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	399	440	EGF-LIKE 6,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	441	480	EGF-LIKE 7,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	481	524	EGF-LIKE 8,	CALCIUM-BINDING	(POTENTIAL)
FT	DOMAIN	525	566	EGF-LIKE 9,	CALCIUM-BINDING	(INCOMPLETE)

FT DISULFID 465 479 BY SIMILARITY.  
 FT DISULFID 485 498 BY SIMILARITY.  
 FT DISULFID 494 507 BY SIMILARITY.  
 FT DISULFID 509 523 BY SIMILARITY.  
 FT DISULFID 529 542 BY SIMILARITY.  
 FT DISULFID 536 551 BY SIMILARITY.  
 FT DISULFID 556 585 BY SIMILARITY.  
 FT CARBOHYD 98 98 POTENTIAL.  
 FT CARBOHYD 447 447 POTENTIAL.  
 FT CARBOHYD 535 535 POTENTIAL.  
 FT CARBOHYD 539 539 POTENTIAL.  
 FT CONFLICT 36 36 C -> S (IN REF. 2).  
 FT CONFLICT 41 42 HR -> SH (IN REF. 2).  
 SQ SEQUENCE 601 AA; 65485 MW; 3A6E492B CRC32;

Query Match 10.1%; Score 96; DB 1; Length 601;  
 Best Local Similarity 57.1%; Pred. No. 1.12e-01;  
 Matches 16; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

Db 10 VPLPILLGGLALLAAGVDADVILEACC 37  
 QY 13 VTLTLLLPALLSLGVDQPLPD-CC 39

RESULT 6  
 ID FBLC\_HUMAN STANDARD; PRT; 683 AA.  
 AC P23144;  
 DT 01-NOV-1991 (REL. 20, CREATED)  
 DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE FIBULIN-1, ISOFORM C PRECURSOR.  
 GN FBLN1.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 91100426.  
 RA ARGRAVES W.S., TRAN H., BURGESS W.H., DICKERSON K.;  
 RT "Fibulin is an extracellular matrix and plasma glycoprotein with  
 repeated domain structure";  
 RL J. CELL BIOL. 111:3155-3164(1990).  
 [2]  
 RP SEQUENCE OF 30-44.  
 RX MEDLINE; 89354537.  
 RA ARGRAVES W.S., DICKERSON K., BURGESS W.H., RUOSLAHTI E.;  
 RT "Fibulin, a novel protein that interacts with the fibronectin  
 receptor beta subunit cytoplasmic domain";  
 RL CELL 58:623-629(1989).  
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.  
 CC -!- ALTERNATIVE PRODUCTS: FOUR FORMS OF FIBULIN-1, A (AC P23142), B  
 (AC P23143), C (SHOWN HERE) AND D (AC P37888); DIFFERING ONLY IN  
 THEIR C-TERMINAL REGIONS, ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- SIMILARITY: CONTAINS 3 ANAPHYLATOXIN-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X53743; G31419; -  
 CC DR PIR; A32826; A32826.  
 CC DR PIR; C36346; C36346.  
 CC DR MIM; 135820; -  
 CC DR PROSITE; PS00010; ASX\_HYDROXYL; 4.  
 CC DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 CC DR PROSITE; PS01177; ANAPHYLATOXIN\_1; 3.  
 CC DR PROSITE; PS01178; ANAPHYLATOXIN\_2; 3.  
 CC DR PROSITE; PS01185; EGF\_2; 3.

DR PROSITE; PS01187; EGF\_CA; 8.  
 DR PFAM; PF00008; EGF; 5.  
 DR HSPF; P35555; IEMO.  
 KW SIGNAL; ALTERNATIVE SPLICING; GLYCOPROTEIN; EXTRACELLULAR MATRIX;  
 REPEAT; PLASMA; EGF-LIKE DOMAIN; CALCIUM-BINDING.  
 FT SIGNAL 1 29  
 FT CHAIN 30 683 FIBULIN-1, ISOFORM C.  
 FT DOMAIN 36 144 3 X ANAPHYLATOXIN REPEATS.  
 FT REPEAT 36 76 ANAPHYLATOXIN-LIKE 1.  
 FT REPEAT 77 111 ANAPHYLATOXIN-LIKE 2.  
 FT REPEAT 112 144 ANAPHYLATOXIN-LIKE 3.  
 FT DOMAIN 176 215 EGF-LIKE 1.  
 FT DOMAIN 216 261 EGF-LIKE 2.  
 FT DOMAIN 262 307 EGF-LIKE 3.  
 FT DOMAIN 308 355 EGF-LIKE 4.  
 FT DOMAIN 356 398 EGF-LIKE 5.  
 FT DOMAIN 399 440 EGF-LIKE 6.  
 FT DOMAIN 441 480 EGF-LIKE 7.  
 FT DOMAIN 481 524 EGF-LIKE 8.  
 FT DOMAIN 525 569 EGF-LIKE 9.  
 FT DISULFID 36 61 BY SIMILARITY.  
 FT DISULFID 37 68 BY SIMILARITY.  
 FT DISULFID 50 69 BY SIMILARITY.  
 FT DISULFID 78 109 BY SIMILARITY.  
 FT DISULFID 91 110 BY SIMILARITY.  
 FT DISULFID 112 136 BY SIMILARITY.  
 FT DISULFID 113 143 BY SIMILARITY.  
 FT DISULFID 126 144 BY SIMILARITY.  
 FT DISULFID 180 190 BY SIMILARITY.  
 FT DISULFID 186 199 BY SIMILARITY.  
 FT DISULFID 201 214 BY SIMILARITY.  
 FT DISULFID 220 233 BY SIMILARITY.  
 FT DISULFID 227 242 BY SIMILARITY.  
 FT DISULFID 248 260 BY SIMILARITY.  
 FT DISULFID 266 279 BY SIMILARITY.  
 FT DISULFID 273 288 BY SIMILARITY.  
 FT DISULFID 294 306 BY SIMILARITY.  
 FT DISULFID 312 325 BY SIMILARITY.  
 FT DISULFID 319 334 BY SIMILARITY.  
 FT DISULFID 341 354 BY SIMILARITY.  
 FT DISULFID 360 373 BY SIMILARITY.  
 FT DISULFID 367 382 BY SIMILARITY.  
 FT DISULFID 384 397 BY SIMILARITY.  
 FT DISULFID 403 415 BY SIMILARITY.  
 FT DISULFID 411 424 BY SIMILARITY.  
 FT DISULFID 426 439 BY SIMILARITY.  
 FT DISULFID 445 454 BY SIMILARITY.  
 FT DISULFID 450 463 BY SIMILARITY.  
 FT DISULFID 465 479 BY SIMILARITY.  
 FT DISULFID 485 498 BY SIMILARITY.  
 FT DISULFID 494 507 BY SIMILARITY.  
 FT DISULFID 509 523 BY SIMILARITY.  
 FT DISULFID 529 542 BY SIMILARITY.  
 FT DISULFID 536 551 BY SIMILARITY.  
 FT DISULFID 556 568 BY SIMILARITY.  
 FT CARBOHYD 98 98 POTENTIAL.  
 FT CARBOHYD 447 447 POTENTIAL.  
 FT CARBOHYD 535 535 POTENTIAL.  
 FT CARBOHYD 539 539 POTENTIAL.  
 FT CONFLICT 36 36 C -> S (IN REF. 2).  
 FT CONFLICT 41 42 HR -> SH (IN REF. 2).  
 SQ SEQUENCE 683 AA; 74475 MW; 67EFD0D6 CRC32;

Query Match 10.1%; Score 96; DB 1; Length 683;  
 Best Local Similarity 57.1%; Pred. No. 1.12e-01;  
 Matches 16; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

Db 10 VPLPILLGGLALLAAGVDADVILEACC 37  
 QY 13 VTLTLLLPALLSLGVDQPLPD-CC 39

RESULT 7



RL PHILOS. TRANS. R. SOC. LOND., B, BIOL. SCI. 306:345-354(1984).  
CC -!- FUNCTION: C1Q ASSOCIATES WITH THE PROENZYME CLR AND C1S TO YIELD  
CC C1, THE FIRST COMPONENT OF THE SERUM COMPLEMENT SYSTEM. THE  
CC COLLAGEN-LIKE REGIONS OF C1Q INTERACT WITH THE CA(2+)-DEPENDENT  
CC CIR(2)C1S(2) PROENZYME COMPLEX, AND EFFICIENT ACTIVATION OF C1  
CC TAKES PLACE ON INTERACTION OF THE GLOBULAR HEADS OF C1Q WITH THE  
CC FC REGIONS OF IGG OR IGM ANTIBODY PRESENT IN IMMUNE COMPLEXES.  
CC -!- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q,  
CC R AND S IN THE MOLAR RATION OF 1:2:2.  
CC -!- SUBUNIT: C1Q SUBCOMPONENT IS COMPOSED OF NINE SUBUNITS, SIX OF  
CC WHICH ARE DISULFIDE-LINKED DIMERS OF THE A & B CHAINS, AND THREE  
CC -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X03084; G573114; -;  
CC EMBL; M36278; G179638; -;  
CC PIR; A03286; C1HUQB.  
CC PIR; B23422; B23422.  
CC MIN; 120570; -;  
CC PROSITE; PS01113; C1Q; 1.  
CC PFAM; PF00386; C1Q; 1.  
CC COMPLEMENT PATHWAY; PLASMA; HYDROXYLATION; GLYCOPROTEIN; COLLAGEN;  
KW REPEAT: SIGNAL.  
FT SIGNAL 1 25  
FT CHAIN 26 251  
FT DOMAIN 29 112  
FT DOMAIN 113 251  
FT MOD\_RES 26 26  
FT DISULFID 29 29  
FT MOD\_RES 33 33  
FT MOD\_RES 36 36  
FT MOD\_RES 39 39  
FT MOD\_RES 42 42  
FT MOD\_RES 51 51  
FT MOD\_RES 54 54  
FT MOD\_RES 57 57  
FT CARBOHYD 57 57  
FT MOD\_RES 60 60  
FT CARBOHYD 60 60  
FT MOD\_RES 63 63  
FT MOD\_RES 75 75  
FT MOD\_RES 81 81  
FT MOD\_RES 84 84  
FT MOD\_RES 90 90  
FT MOD\_RES 96 96  
FT CARBOHYD 96 96  
FT MOD\_RES 99 99  
FT MOD\_RES 102 102  
FT MOD\_RES 105 105  
FT MOD\_RES 108 108  
FT CARBOHYD 108 108  
FT VARIANT 174 174  
FT CONFLICT 26 26  
FT CONFLICT 83 83  
FT CONFLICT 98 98  
SQ SEQUENCE 251 AA; 26459 MW; 23D5197F CRC32;  
Query Match 9.5%; Score 91; DB 1; Length 251;  
Best Local Similarity 53.3%; Pred. No. 5.61e-01;  
Matches 8; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 2 KIPWGSIPVLMLLL 16

I:|:|:|:|:|:|

Qy 7 KVPWAAVTLMLLLL 21

RESULT 11  
ID LCAT\_RAT STANDARD; PRT; 440 AA.  
AC P18424;  
DT 01-NOV-1990 (REL. 16, CREATED)  
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)  
DE PHOSPHATIDYLCHOLINE-STEROL ACYLTRANSFERASE PRECURSOR (EC 2.3.1.43)  
DE (LECITHIN-CHOLESTEROL ACYLTRANSFERASE) (PHOSPHOLIPID-CHOLESTEROL  
DE ACYLTRANSFERASE).  
GN LCAT.  
OS RATTUS NORVEGICUS (RAT)  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 90384859.  
RA MERONI G., MARGARETTI N., MAGNAGHI P., TARAMELLI R.;  
RT "Nucleotide sequence of the cDNA for lecithin-cholesterol acyl  
RT transferase (LCAT) from the rat.";  
RL NUCLEIC ACIDS RES. 18:5308-5308(1990).  
CC -!- FUNCTION: CENTRAL ENZYME IN THE EXTRACELLULAR METABOLISM OF PLASMA  
CC LIPOPROTEINS. AMONG OTHER SUBSTRATES IT ESTERIFIES THE FREE  
CC CHOLESTEROL TRANSPORTED IN PLASMA LIPOPROTEINS.  
CC -!- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + STEROL = STEROL ESTER +  
CC 1-ACYLGLYCEROPHOSPHOCHOLINE (PALMITOYL, OLEOYL, AND LINOLEOYL CAN  
CC BE TRANSFERRED; A NUMBER OF STEROLS, INCLUDING CHOLESTEROL, CAN  
CC ACT AS ACCEPTOR).  
CC -!- ENZYME REGULATION: APOLIPOPROTEIN A-I IS A POTENT ACTIVATOR FOR  
CC THIS ENZYME.  
CC -!- SIMILARITY: PARTIAL WITH LIPASES (PANCREATIC, GASTRIC, HEPATIC,  
CC LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X54096; G56564; -;  
CC PIR; S11214; XXRTN.  
CC PROSITE; PS00120; LIPASE SER; 1.  
KW TRANSFERASE; ACYLTRANSFERASE; LIPID METABOLISM; GLYCOPROTEIN; SIGNAL.  
FT SIGNAL 1 24  
FT CHAIN 25 440  
FT ACT\_SITE 205 205  
FT DISULFID 74 98  
FT DISULFID 337 380  
FT CARBOHYD 44 44  
FT CARBOHYD 108 108  
FT CARBOHYD 296 296  
FT CARBOHYD 397 397  
FT CARBOHYD 408 408  
SQ SEQUENCE 440 AA; 49727 MW; 5DEB1ECD CRC32;  
Query Match 9.5%; Score 91; DB 1; Length 440;  
Best Local Similarity 62.5%; Pred. No. 5.61e-01;  
Matches 15; Conservative 2; Mismatches 5; Indels 2; Gaps 1;  
Db 1 MGLPGS--PWQWVLLLLGLLLPPA 22  
I:|:|:|:|:|:|  
Qy 1 MNLSTKVPWAAVTLMLLLLPPA 24  
RESULT 12  
ID LCAT\_HUMAN STANDARD; PRT; 440 AA.  
AC P04180;  
DT 20-MAR-1987 (REL. 04, CREATED)  
DT 20-MAR-1987 (REL. 04, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE PHOSPHATIDYLCHOLINE-STEROL ACYLTRANSFERASE PRECURSOR (EC 2.3.1.43)

RT "Complete deficiency of plasma lecithin-cholesterol acyltransferase  
 RT (LCAT) activity due to a novel homozygous mutation (Gly-30-Ser) in  
 RT the LCAT gene.";  
 RL HUM. MUTAT. 8:79-82(1996).  
 CC -!- FUNCTION: CENTRAL ENZYME IN THE EXTRACELLULAR METABOLISM OF PLASMA  
 CC LIPOPROTEINS. AMONG OTHER SUBSTRATES IT ESTERIFIES THE FREE  
 CC CHOLESTEROL TRANSPORTED IN PLASMA LIPOPROTEINS.  
 CC -!- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + STEROL = STEROL ESTER +  
 CC 1-ACYLGLYCEROPHOSPHOCHOLINE (PALMITOYL, OLEOYL, AND LINOLEOYL CAN  
 CC BE TRANSFERRED; A NUMBER OF STEROLS, INCLUDING CHOLESTEROL, CAN  
 CC ACT AS ACCEPTOR).  
 CC -!- ENZYME REGULATION: APOLIPOPROTEIN A-I IS A POTENT ACTIVATOR FOR  
 CC THIS ENZYME.  
 CC -!- DISEASE: DEFECTS IN LCAT ARE THE CAUSE OF NORUM AND FISH EYE  
 CC DISEASES.  
 CC -!- SIMILARITY: PARTIAL WITH LIPASES (PANCREATIC, GASTRIC, HEPATIC,  
 CC LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M12625; G307117; -;  
 DR EMBL: X06537; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL: M26268; G187025; -;  
 DR EMBL: X04981; G34287; -;  
 DR EMBL: M17959; G386858; -;  
 DR PIR: A00571; XXHUN.  
 DR PIR: A29661; A29661.  
 DR PIR: A25575; A25575.  
 DR PIR: JQ0036; JQ0036.  
 DR MIM: 136120; -;  
 DR MIM: 245900; -;  
 DR PROSITE: PS00120; LIPASE\_SER; 1.  
 DR TRANSFERASE; ACYLTRANSFERASE; LIPID METABOLISM; GLYCOPROTEIN; SIGNAL;  
 KW POLYMORPHISM; DISEASE MUTATION.  
 FT SIGNAL 1 24  
 FT CHAIN 25 440 LECITHIN-CHOLESTEROL ACYLTRANSFERASE.  
 FT ACT\_SITE 205 205 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 74 98  
 FT DISULFID 337 380  
 FT CARBOHYD 44 44  
 FT CARBOHYD 108 108  
 FT CARBOHYD 296 296

...  
 Note: remainder of annotations omitted.

Query Match 9.5%; Score 91; DB 1; Length 440;  
 Best Local Similarity 81.3%; Pred. No. 5.61e-01;  
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 7 PWQWVTLGLGLPPA 22  
 || ||||| |||||  
 QY 9 PWAATVLLLLLLPPA 24

RESULT 13  
 ID PDGB\_HUMAN STANDARD; PRT; 241 AA.  
 AC P01127; P78431;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE PLATELET-DERIVED GROWTH FACTOR, B CHAIN PRECURSOR (PDGF B-CHAIN)  
 DE (PDGF-2) (BACAPLERMIN).  
 GN PDGFB OR C-SIS.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE; 84250225.  
 RA JOSEPHS S.F., RATNER L., CLARKE M.F., WESTIN E.H., REITZ M.S.,  
 RA WONG-STAAL F.;  
 RT "Transforming potential of human c-sis nucleotide sequences encoding  
 RT platelet-derived growth factor.";  
 RL SCIENCE 225:636-639(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 86205961.  
 RA RAO C.D., IGARASHI H., CHIU I.-M., ROBBINS K.C., AARONSON S.A.;  
 RT "Structure and sequence of the human c-sis/platelet-derived growth  
 RT factor 2 (SIS/PDGF2) transcriptional unit.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 83:2392-2396(1986).  
 RN [3]  
 RP SEQUENCE OF 22-241 FROM N.A.  
 RX MEDLINE; 84205633.  
 RA CHIU I.-M., REDDY E.P., GIVOL D., ROBBINS K.C., TRONICK S.R.,  
 RA AARONSON S.A.;  
 RT "Nucleotide sequence analysis identifies the human c-sis  
 RT proto-oncogene as a structural gene for platelet-derived growth  
 RT factor.";  
 RL CELL 37:123-129(1984).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 85296313.  
 RA COLLINS T., GINSBURG D., BOSS J.M., ORKIN S.H., POBER J.S.;  
 RT "Cultured human endothelial cells express platelet-derived growth  
 RT factor B chain: cDNA cloning and structural analysis.";  
 RL NATURE 316:748-750(1985).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 85269623.  
 RA RATNER L., JOSEPHS S.F., JARRETT R., REITZ M.S., WONG-STAAL F.;  
 RT "Nucleotide sequence of transforming human c-sis cDNA clones with  
 RT homology to platelet-derived growth factor.";  
 RL NUCLEIC ACIDS RES. 13:5007-5018(1985).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 87217119.  
 RA RAO C.D., IGARASHI H., PECH M.W., ROBBINS K.C., AARONSON S.A.;  
 RT "Oncogenic potential of the human platelet-derived growth factor  
 RT transcriptional unit.";  
 RL COLD SPRING HARB. SYMP. QUANT. BIOL. 51:959-966(1986).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RX BURGESS J., ODELL C.;  
 RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [8]  
 RP SEQUENCE OF 1-53 FROM N.A.  
 RX MEDLINE; 97141927.  
 RA SIMON M.-P., PEDEUTOUR F., SIRVENT N., GROSSEGEORGE J., MINOLETTI F.,  
 RA COINDRE J.-M., TERRIER-LACOMBE M.-J., MANDAH N., CRAVER R.D.,  
 RA BLIN N., SOZZI G., TURC-CAREL C., O'BRIEN K.P., KEDRA D.,  
 RA FRANSSON I., GUILBAUD C., DUMANSKI J.P.;  
 RT "Deregulation of the platelet-derived growth factor B-chain gene via  
 RT fusion with collagen gene COL1A1 in dermatofibrosarcoma protuberans  
 RT and giant-cell fibroblastoma.";  
 RL NAT. GENET. 15:95-98(1997).  
 RN [9]  
 RP SEQUENCE OF 26-241 FROM N.A.  
 RX MEDLINE; 86164981.  
 RA WEICH H.A., SEBALD W., SCHAIER H.U., HOPPE J.;  
 RT "The human osteosarcoma cell line U-2 OS expresses a 3.8 kilobase  
 RT mRNA which codes for the sequence of the PDGF-B chain.";  
 RL FEBS LETT. 198:344-348(1986).  
 RN [10]  
 RP SEQUENCE OF 153-200 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE; 84236121.  
 RA JOHNSON A., HELDIN C.H., WASTESON A., WESTERMARK B., DEUEL T.F.,  
 RA HUANG J.S., SEEBURG P.H., GRAY A., ULLRICH A., SCRACE G.,  
 RA STROOBANT P., WATERFIELD M.D.;  
 RT "The c-sis gene encodes a precursor of the B chain of

[illegible]

ARCH. BIOCHEM. BIOPHYS. 307:361-368(1993).

[2]

SEQUENCE FROM N.A.

STRAIN-FISHER 344; TISSUE=TRACHEA;

MEDLINE: 95168876;

KITZLER J., HILL E., HARDMAN R., REDDY N., PHILPOT R., ELING T.E.;

"Analysis and quantitation of splicing variants of the TPA-inducible PGH-1 mRNA in rat tracheal epithelial cells.;"

ARCH. BIOCHEM. BIOPHYS. 316:856-863(1995).

CC !- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN REGULATING OR PROMOTING CELL PROLIFERATION IN SOME NORMAL AND NEOPLASTICALLY TRANSFORMED CELLS.

CC !- CATALYTIC ACTIVITY: ARACHIDONATE + AH(2) + 2, O(2) = PROSTAGLANDIN H2 + A + H(2)O.

CC !- PATHWAY: FIRST STEP IN THE FORMATION OF PROSTAGLANDINS AND THROMBOXANES.

CC !- SUBUNIT: HOMODIMER (BY SIMILARITY).

CC !- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MICROSOMAL MEMBRANE.

CC !- THIS ENZYME ACTS BOTH AS A DIOXYGENASE AND AS A PEROXIDASE.

CC !- THIS ENZYME IS THE TARGET OF NONSTEROIDAL ANTI-INFLAMMATORY DRUGS SUCH AS ASPIRIN.

CC !- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.

CC !- SIMILARITY: BELONGS TO THE PROSTAGLANDIN G/H SYNTHASE FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL: U03388; G415638; -

DR EMBL: S67721; G450556; -

DR EMBL: U18060; G603052; -

DR PROSITE: PS00022; EGF\_1; FALSE\_NEG.

DR PROSITE: PS01186; EGF\_2; FALSE\_NEG.

DR PFAM: PF00008; EGF\_1.

KW OXIDOREDUCTASE; DIOXYGENASE; PEROXIDASE; GLYCOPROTEIN; ACETYLATION;

KW PROSTAGLANDIN BIOSYNTHESIS; HEME; IRON; SIGNAL; MEMBRANE;

KW EGF-LIKE DOMAIN.

SIGNAL	1	26	BY SIMILARITY.
CHAIN	27	602	PROSTAGLANDIN G/H SYNTHASE 1.
DOMAIN	34	72	EGF-LIKE.
ACT_SITE	209	209	DISTAL HISTIDINE (BY SIMILARITY).
ACT_SITE	387	387	CYCLOOXYGENASE (BY SIMILARITY).
BINDING	390	390	PROXIMAL HEME LIGAND (BY SIMILARITY).
MOD_RES	532	532	ASPIRIN-ACETYLATED SERINE.
DISULFID	38	49	BY SIMILARITY.
DISULFID	43	59	BY SIMILARITY.
DISULFID	61	71	BY SIMILARITY.
DISULFID	39	161	BY SIMILARITY.
DISULFID	571	577	BY SIMILARITY.
CARBOHYD	70	70	POTENTIAL.
CARBOHYD	106	106	POTENTIAL.
CARBOHYD	146	146	POTENTIAL.
CARBOHYD	412	412	POTENTIAL.
CONFLICT	36	36	N -> I (IN REF. 1).
CONFLICT	116	117	RL -> GW (IN REF. 1).
CONFLICT	119	119	I -> L (IN REF. 1).
CONFLICT	192	192	G -> A (IN REF. 1).
CONFLICT	263	263	V -> L (IN REF. 1).
CONFLICT	274	274	L -> K (IN REF. 1).
CONFLICT	290	290	G -> A (IN REF. 1).
CONFLICT	339	339	I -> R (IN G450556).
CONFLICT	344	344	K -> E (IN REF. 1).
CONFLICT	381	381	L -> M (IN REF. 1).
CONFLICT	392	392	L -> F (IN REF. 1).
SEQUENCE	602 AA;	69032 MW;	7E3888D7 CRC32;

Query Match 9.3%; Score 89; DB 1; Length 602;

Best Local Similarity 40.0%;

Matches 16; Conservative 7; Mismatches 15; Indels 2; Gaps 2;

Query Match 9.3%; Score 89; DB 1; Length 602;  
Best Local Similarity 40.0%; Pred. No. 1.05e+00;  
Matches 16; Conservative 7; Mismatches 15; Indels





```

DATE          30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
              05-Sep-1997
ACCESSIONS    A34158; S21370
REFERENCE      Warden, C.H.; Langner, C.A.; Gordon, J.I.; Taylor, B.A.;
#authors      McLean, J.W.; Lusis, A.J.
#journal      J. Biol. Chem. (1989) 264:21573-21581
#title        Tissue-specific expression, developmental regulation, and
              chromosomal mapping of the lecithin:cholesterol
              acyltransferase gene. Evidence for expression in brain and
              testes as well as liver.
#cross-references MUID:90094326
#accession     A34158
#molecule_type mRNA
#residues      1-438 #label WAR
#cross-references GB:J05154; NID:g198759; PID:g293697
#note          the authors translated the codon ATG for residue 411 as
              Leu
REFERENCE      S21370
#authors      Meroni, G.; Malfaretti, N.; Magnaghi, P.; Taramelli, R.
#submission   submitted to the EMBL Data Library, July 1990
#description   Promoter and 5' flanking sequences of the mouse LCAT gene.
#accession     S21370
#molecule_type DNA
#residues      1-14 #label MER
#cross-references EMBL:g54095; NID:g52873; PID:g52874
COMMENT        The active enzyme catalyzes the transfer of acyl groups from
              lecithin to sterol to form sterol esters. Palmitoyl, oleoyl, and
              linoleoyl residues can be transferred; a number of sterols,
              including cholesterol, can act as acceptor. Apolipoprotein A-I is
              a potent activator for this enzyme.
CLASSIFICATION #superfamily phosphatidylcholine--sterol acyltransferase
KEYWORDS        acyltransferase; glycoprotein; lipid metabolism; lipoprotein
FEATURE         1-24
              25-438
              44,108,296,408
              #length 438 #molecular-weight 49765 #checksum 1794
SUMMARY          Query Match          10.1%; Score 96; DB 1; Length 438;
              Best Local Similarity 62.5%; Pred. No. 4.07e-01;
              Matches 15; Conservative 2; Mismatches 5; Indels 2; Gaps 1;

Db 1 MGLPGS--PWQRYLLGLLLPPA 22
  ||||| || ||||| |||||
QY 1 MNLSTKVPWAATLGLLLPPA 24

RESULT 6
ENTRY      B36346 #type complete
TITLE      fibulin 1 precursor, splice form B - human
ORGANISM   Homo sapiens #common_name man
#formal_name Homo sapiens #common_name man
#accession U00000
#molecule_type mRNA
#residues 1-601 #label ARG
#cross-references GB:X53742; NID:g31416; PID:g31417
#note      GDB:FBLN1; FBLN
#cross-references GDB:278285; OMIM:135820
#map_position 22q13.3-22q13.3
#classification #superfamily EGF homology
KEYWORDS    alternative splicing

```

```

FEATURE        #domain EGF homology #label EGF
485-523
SUMMARY        #length 601 #molecular-weight 65485 #checksum 6896

Query Match    10.1%; Score 96; DB 2; Length 601;
Best Local Similarity 57.1%; Pred. No. 4.07e-01;
Matches 16; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

Db 10 VPLPLLLLGGLAAGVDADVLEAAC 37
  ||||| ||||| ||||| |||||
QY 13 VTLLLLLLPPALLSLGVDAQPLD-CC 39

RESULT 7
ENTRY      C36346 #type complete
TITLE      fibulin 1 precursor, splice form C - human
ALTERNATE_NAMES fibulin C
CONTAINS      fibulin 1 splice form A; fibulin 1 splice form C
ORGANISM      Homo sapiens #common_name man
#formal_name Homo sapiens #common_name man
#accession U00000
#molecule_type mRNA
#residues 1-683 #label ARG
#cross-references GB:X53743; NID:g31418; PID:g31419
#note      GDB:FBLN1; FBLN
#cross-references GDB:278285; OMIM:135820
#map_position 22q13.3-22q13.3
#classification #superfamily EGF homology
KEYWORDS    alternative splicing; glycoprotein
FEATURE      1-29
              30-683
              98,535,539
              #length 683 #molecular-weight 74475 #checksum 7443
SUMMARY          Query Match          10.1%; Score 96; DB 2; Length 683;
              Best Local Similarity 57.1%; Pred. No. 4.07e-01;
              Matches 16; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

Db 10 VPLPLLLLGGLAAGVDADVLEAAC 37
  ||||| ||||| ||||| |||||
QY 13 VTLLLLLLPPALLSLGVDAQPLD-CC 39

RESULT 8
ENTRY      S28774 #type fragment
TITLE      collagen alpha chain - tube worm (Riftia pachyptila)
              (fragment)
ORGANISM      Riftia pachyptila
#formal_name Riftia pachyptila
#accession U00000
#molecule_type mRNA
#residues 1-601 #label ARG
#cross-references GB:X53742; NID:g31416; PID:g31417
#note      GDB:FBLN1; FBLN
#cross-references GDB:278285; OMIM:135820
#map_position 22q13.3-22q13.3
#classification #superfamily EGF homology
KEYWORDS    alternative splicing

```

```

GENETICS
#gene GDB:NPR1; NPRA: ANPRA
#map_position lq21-lq22
CLASSIFICATION #superfamily membrane-bound guanylate cyclase; guanylate
cyclase catalytic domain homology; natriuretic
peptide-binding domain homology; protein kinase homology
KEYWORDS ATP; carbon-oxygen lyase; cGMP biosynthesis; glycoprotein;
hormone receptor; phosphorus-oxygen lyase; transmembrane
protein
FEATURE
1-32 #domain signal sequence #status predicted #label SIG\
33-1061 #product natriuretic peptide receptor A #status
predicted #label MAT\
33-473 #domain extracellular #status predicted #label EXT\
113-453 #domain natriuretic peptide-binding domain homology
#label NPB\
474-494 #domain transmembrane #status predicted #label TM\
495-1061 #domain intracellular #status predicted #label INT\
526-808 #domain protein kinase homology #label KIN\
828-1055 #domain guanylate cyclase catalytic domain homology
#label CAT\
34,45,212,338,379, #binding_site carbohydrate (Asn) (covalent) #status
386,427 predicted
SUMMARY #length 1061 #molecular-weight 118918 #checksum 6297
Query Match 9.7%; Score 93; DB 1; Length 1061;
Best Local Similarity 48.6%; Pred. No. 9,78e-01;
Matches 1; Conservative 6; Mismatches 10; Indels 2; Gaps 2;
Db 1 MPGPRRPGSRRLRLLLLPPLLLLRGSHAGNL 35
: : : : :
Qy 3 LPSTKVPWAA-VTLRLLLLPALLSL-GVDAQPL 35
: : : : :

RESULT 10
ENTRY ClHQB #type complete
TITLE complement subcomponent Clq chain B precursor - human
ALTERNATE_NAMES complement subcomponent Clq beta chain
ORGANISM #formal_name Homo sapiens #common_name man
DATE 22-May-1981 #sequence_revision 31-May-1996 #text_change
22-May-1998
ACCESSIONS B23422; A23422; B90304; A90301; B90315; A03206
REFERENCE Reid, K.B.M.
Biochem. J. (1985) 231:729-735
Molecular cloning and characterization of the complementary
DNA and gene coding for the B-chain of subcomponent Clq of
the human complement system.
#cross-references MUID:86076906
#accession B23422
#molecule_type DNA
#residues 'HS',1-32 #label REI
#note the authors translated the codon AGT for the second
position as Arg; they were uncertain about the
location of the initiation codon
#accession A23422
#molecule_type mRNA
#residues 28-253 #label REI
#cross-references EMBL:X03084
#note the authors translated the codon ACA for residue 46 as
Ile
REFERENCE A90304
#authors Reid, K.B.M.
#journal Biochem. J. (1979) 179:367-371
#title Complete amino acid sequences of the three collagen-like
regions present in subcomponent Clq of the first component
of human complement.
#cross-references MUID:80020137
#accession B90304
#molecule_type protein

```

```

#residues 'E',29-84,'D',86-99,'P',101-135 #label RE5
REFERENCE A90301
#authors Reid, K.B.M.; Thompson, E.O.P.
#journal Biochem. J. (1978) 173:863-868
#title Amino acid sequence of the N-terminal 108 amino acid residues
of the B chain of subcomponent Clq of the first component
of human complement.
#cross-references MUID:79041552
#accession A90301
#molecule_type protein
#residues 28-99,'P',101-195 #label RE3
REFERENCE A90315
#authors Reid, K.B.M.; Gagnon, J.; Frampton, J.
#journal Biochem. J. (1982) 203:559-569
#title Completion of the amino acid sequences of the A and B chains
of subcomponent Clq of the first component of human
complement.
#cross-references MUID:82283890
#accession B90315
#molecule_type protein
#residues 136-253 #label RE4
#note 176-Clx may also be present
COMMENT The first component of complement is a calcium-dependent complex of
the three subcomponents Clq,Clr, and C1s. Subcomponent Clq binds
to immunoglobulin complexes, with resulting serial activation of
Clr (enzyme), C1s (proenzyme), and the other eight components of
complement.
The Clq subcomponent is composed of nine subunits, six of which are
disulfide-linked dimers of the A (see PIR:CIHUQA) and B chains,
and three of which are disulfide-linked dimers of the C (see
PIR:CIHUQC) chain. Equimolar amounts of the A, B, and C chains
are found after reduction of the disulfide bonds.
GENETICS
#gene GDB:CIQB
#cross-references GDB:119043; OMIM:120570
#map_position lp36.3-lp34.1
CLASSIFICATION #superfamily complement subcomponent Clq chain A; complement
Clq carboxyl-terminal homology
KEYWORDS complement pathway; glycoprotein; heterodimer; hydroxylysine;
hydroxyproline; plasma; pyroglutamic acid; triple helix
FEATURE
1-27 #domain signal sequence #status predicted #label SIG\
28-253 #product complement subcomponent Clq chain B #status
experimental #label MAT\
33-116 #domain collagenous, triple helix #label COL\
123-249 #domain complement Clq carboxyl-terminal homology #label
Clq\
28 #modified_site pyrrolidone carboxylic acid (Gln) (in
mature form) #status experimental\
31 #disulfide_bonds interchain (to chain A-26) #status
experimental\
35,38,41,53,56,65, #modified_site 4-hydroxyproline (Pro) #status
83,86,101,104,107 experimental\
59,62,77,92,98, #modified_site 5-hydroxylysine (Lys) #status
110 experimental\
59,62,98,110 #binding_site carbohydrate (Lys) (covalent) #status
experimental\
SUMMARY #length 253 #molecular-weight 26722 #checksum 7399
Query Match 9.5%; Score 91; DB 1; Length 253;
Best Local Similarity 53.3%; Pred. No. 1.74e+00;
Matches 8; Conservative 6; Mismatches 1; Indels 0; Gaps 0;
Db 4 KIPWGSIPVLMLLLL 18
: : : : :
Qy 7 KVPWAAVTLMLLLLL 21
: : : : :
RESULT 11
ENTRY XXRTN #type complete
TITLE phosphatidylcholine--sterol O-acyltransferase (EC 2.3.1.43)

```

```

COMMENT  Apolipoprotein A-I (see PIR:LPHUA1) is a potent activator of this
          enzyme.
GENETICS
#gene      GDB:LCAT
##cross-references GDB:119359; OMIM:245900
#map_position 16q22.1-16q22.1
#description catalyzes the transfer of sn-2 fatty acyl groups from
              phosphatidylcholine (lecithin) to sterol to form sterol
              fatty esters and 1-acylglycerophosphocholine
#note      palmitoyl, oleoyl, and linoleoyl residues can be transferred;
              a number of sterols, including cholesterol, can act as
              acceptor.
CLASSIFICATION #superfamily phosphatidylcholine--sterol acyltransferase
KEYWORDS      acyltransferase; cholesterol; glycoprotein; lipid metabolism;
              lipoprotein
FEATURE
1-24         #domain signal sequence #status predicted #label SIG\
25-440       #product phosphatidylcholine--sterol acyltransferase
              #status experimental #label MAT\
202-207      #region interfacial lipid recognition (GXSGXG) motif\
44,108,296,408 #binding_site carbohydrate (Asn) (covalent) #status
              experimental\
431          #binding_site carbohydrate (Thr) (covalent) #status
              experimental\
433          #binding_site carbohydrate (Ser) (covalent) #status
              experimental\
SUMMARY      #length 440 #molecular-weight 49578 #checksum 8696

Query Match      9.5%; Score 91; DB 1; Length 440;
Best Local Similarity 81.3%; Pred. No. 1.74e+00;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 7 PQQWVTLGLGLLPPA 22
   || ||||| |||||
QY 9 PWAATVTLGLLGLLPPA 24

RESULT 13
ENTRY   S39782 #type complete
TITLE   cyclooxygenase 1 - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE    20-May-1994 #sequence_revision 10-Nov-1995 #text_change
        29-Jan-1999
ACCESSIONS S39782
REFERENCE  S39782
#authors   Feng, L.; Sun, W.; Xia, Y.; Tang, W.W.; Chanmugam, P.;
           Soyoola, E.; Wilson, C.B.; Hwang, D.
#journal   Arch. Biochem. Biophys. (1993) 307:361-368
#title     Cloning two isoforms of rat cyclooxygenase: differential
           regulation of their expression.
#cross-references MUID:94099619
#accession S39782
#molecule_type mRNA
#residues 1-602 #label FEN
##cross-references GB:S67721; NID:g460555; PID:g460556
#length 602 #molecular-weight 69168 #checksum 6900

SUMMARY

Query Match      9.5%; Score 91; DB 2; Length 602;
Best Local Similarity 40.0%; Pred. No. 1.74e+00;
Matches 16; Conservative 7; Mismatches 15; Indels 2; Gaps 2;

Db 1 MSRRSLQFPLLLLLLLLLPPVLLTDAGVPSVIP-CC 39
   | : : : ||||| || || : : ||
QY 1 MNLPSKVPWAAVTLGLLGLLPPALLS-LGVDAQPLPDC 39

RESULT 14
ENTRY   S58383 #type complete
TITLE   hypothetical protein 2 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE    12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change
        02-Jul-1998

```

```

ACCESSIONS S58383
REFERENCE  S58382
#authors   Dirks, R.P.H.; Onnekink, C.; Jansen, H.J.; de Jong, A.;
           Bloemers, H.P.J.
#journal   Nucleic Acids Res. (1995) 23:2815-2822
#title     A novel human c-sis mRNA species is transcribed from a
           promoter in c-sis intron 1 and contains the code for an
           alternative PDGF B-like protein.
#accession S58383
#status    preliminary
#molecule_type mRNA
#residues 1-185 #label DIR
#cross-references EMBL:X83705; NID:g951023; PID:g951025
CLASSIFICATION #superfamily platelet-derived growth factor
SUMMARY      #length 185 #molecular-weight 20774 #checksum 2728

Query Match      9.3%; Score 89; DB 2; Length 185;
Best Local Similarity 43.6%; Pred. No. 3.06e+00;
Matches 17; Conservative 11; Mismatches 6; Indels 5; Gaps 5;

Db 29 LQRLHGDPEEDGAELDLNMTSRSHSGGELESIA-RGRR 66
   |||||:: | : | : | : | : ||| : |||
QY 79 LQRLQAN-GNH-AGGI-LTMGR-RAGAELEPYPCGRR 113

RESULT 15
ENTRY   A55030 #type fragment
TITLE   platelet-derived growth factor chain B precursor - human
           (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE    18-Nov-1994 #sequence_revision 18-Nov-1994 #text_change
        17-Mar-1999
ACCESSIONS A55030
REFERENCE  A55030
#authors   Johansson, A.; Heldin, C.H.; Wasteson, A.; Westermark, B.;
           Deuel, T.F.; Huang, J.S.; Seeburg, P.H.; Gray, A.; Ullrich,
           A.; Scrcce, G.; Stroobant, P.; Waterfield, M.D.
#journal   EMBO J. (1984) 3:921-928
#title     The c-sis gene encodes a precursor of the B chain of
           platelet-derived growth factor.
#cross-references MUID:84236121
#accession A55030
#status    preliminary
#molecule_type DNA
#residues 1-230 #label JOH
#cross-references GB:X00556; GB:X00559; GB:X00560; GB:X00561; GB:X00562
CLASSIFICATION #superfamily platelet-derived growth factor
SUMMARY      #length 230 #checksum 3580

Query Match      9.3%; Score 89; DB 2; Length 230;
Best Local Similarity 43.6%; Pred. No. 3.06e+00;
Matches 17; Conservative 11; Mismatches 6; Indels 5; Gaps 5;

Db 33 LQRLHGDPEEDGAELDLNMTSRSHSGGELESIA-RGRR 70
   |||||:: | : | : | : | : ||| : |||
QY 79 LQRLQAN-GNH-AGGI-LTMGR-RAGAELEPYPCGRR 113

Search completed: Fri Aug 20 21:01:34 1999
Job time : 22 secs.

```

(TM)

Result	Query				ID	Description	Pred. No.
	No.	Score	Match	Length			
1	92	9.6	220	4	5175255-4	Patent No. 5175255.	7.65e+00
2	89	9.3	241	3	PCT-US96-0	Sequence 9, Applicatio	1.26e+01
3	89	9.3	241	4	5194596-15	Patent No. 5194596.	1.26e+01
4	89	9.3	241	1	US-08-387-	Sequence 4, Applicatio	1.26e+01
5	88	9.2	254	1	US-08-236-	Sequence 4, Applicatio	1.49e+01
6	87	9.1	226	4	5498500-2	Patent No. 5498600.	1.76e+01
7	87	9.1	241	4	5175255-2	Patent No. 5175255.	1.76e+01
8	87	9.1	241	4	5175255-8	Patent No. 5175255.	1.76e+01
9	87	9.1	282	1	US-08-445-	Sequence 1, Applicatio	1.76e+01
10	85	8.9	241	1	US-08-469-	Sequence 13, Applicati	2.44e+01
11	85	8.9	613	3	PCT-US94-1	Sequence 2, Applicatio	2.44e+01
12	85	8.9	613	2	US-08-465-	Sequence 2, Applicatio	2.44e+01
13	85	8.9	954	2	US-08-749-	Sequence 3, Applicatio	2.44e+01
14	84	8.8	234	1	US-08-299-	Sequence 5, Applicatio	2.88e+01
15	84	8.8	238	1	US-08-453-	Sequence 2, Applicatio	2.88e+01
16	84	8.8	238	1	US-08-240-	Sequence 2, Applicatio	2.88e+01
17	84	8.8	1239	1	US-08-026-	Sequence 3, Applicatio	2.88e+01
18	83	8.7	235	1	US-08-243-	Sequence 6, Applicatio	3.99e+01
19	83	8.7	235	3	PCT-US94-0	Sequence 6, Applicatio	3.99e+01
20	83	8.7	235	2	US-08-993-	Sequence 6, Applicatio	3.99e+01
21	82	8.6	241	4	5219739-15	Patent No. 5219739.	3.99e+01
22	80	8.4	492	1	US-07-794-	Sequence 4, Applicatio	5.50e+01
23	80	8.4	492	1	US-08-001-	Sequence 4, Applicatio	5.50e+01

DE  
Sequence 1, Application US/08445847A  
YY

```
Db 1 MRAPGALLARMSRLLLLLLLLLVKSASSALGVAPASRNETCLGESCA 45
      | |:: :: : ||||| : ||| : : : | :| :
```

AP

\*\*\*\*\*

W P E R L A H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:06:48 1999; Maspar time 3.32 Seconds  
Tabular output not generated. 280.825 Million cell updates/sec

Title: >US-08-938-548B-8  
Description: (1-33) from US08938548B.pep  
Perfect Score: 256  
Sequence: 1 QPLPDCCKRQCTCRLYELHGAGNHAAGILTL 33

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 31.702; Variance 45.079; scale 0.703

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	75	29.3	373	1	ICBW_MOUSE	CASPASE-11 PRECURSOR (	5.93e-02
2	73	28.5	260	1	NMA_HUMAN	PUTATIVE TRANSMEMBRANE	1.38e-01
3	72	28.1	815	1	GYRB_MYXXA	DNA GYRASE SUBUNIT B (	2.10e-01
4	71	27.7	413	1	PPAW_CAEEL	PUTATIVE ACID PHOSPHAT	3.18e-01
5	70	27.3	214	1	R10A_TRYBR	60S RIBOSOMAL PROTEIN	4.80e-01
6	69	27.0	197	1	MCS_MOUSE	SPERM MITOCHONDRIAL CA	7.21e-01
7	69	27.0	461	1	YUL2_CAEEL	PUTATIVE FORKHEAD-RELA	7.21e-01
8	68	26.6	641	1	TETS_LISMO	TETRACYCLINE RESISTANC	1.08e+00
9	68	26.6	646	1	TETS_LACLA	TETRACYCLINE RESISTANC	1.08e+00
10	67	26.2	299	1	ALC_RABIT	IG ALPHA CHAIN C REGIO	1.61e+00
11	67	26.2	923	1	RE11_SCHPO	MEIOTIC RECOMBINATION	1.61e+00
12	66	25.8	347	1	GALE_RAT	UDP-GLUCOSE 4-EPIMERAS	2.38e+00
13	66	25.8	348	1	GALE_HUMAN	UDP-GLUCOSE 4-EPIMERAS	2.38e+00
14	66	25.8	360	1	PURK_FSEAE	PHOSPHORIBOSYLAMINOIM	2.38e+00
15	66	25.8	380	1	LEU3_PHACH	3-ISOPROPYLMALATE DEHY	2.38e+00
16	66	25.8	425	1	IF15_MOUSE	INTERFERON-ACTIVATABLE	2.38e+00
17	65	25.4	155	1	Y115_METJA	HYPOTHETICAL PROTEIN M	3.52e+00
18	65	25.4	273	1	MD12_SCHPO	MITOCHONDRIAL INHERITA	3.52e+00
19	64	25.0	76	1	TX03_AGEAP	OMEGA-AGATOXIN IIIA.	5.17e+00
20	64	25.0	83	1	TX4B_AGEAP	OMEGA-AGATOXIN IVB PRE	5.17e+00
21	64	25.0	103	1	BOIA_HAEIN	BOLA PROTEIN HOMOLOG.	5.17e+00
22	64	25.0	125	1	AGSW_VULVU	AGOUTI SWITCH PROTEIN	5.17e+00
23	64	25.0	131	1	AGSW_MOUSE	AGOUTI SWITCH PROTEIN	5.17e+00

24	64	25.0	132	1	AGSW_HUMAN	AGOUTI SWITCH PROTEIN	5.17e+00
25	64	25.0	397	1	CBAC_ALCSP	1-CARBOXY-3-CHLORO-3,4	5.17e+00
26	64	25.0	493	1	VPE_VICSA	VACUOLAR PROCESSING EN	5.17e+00
27	63	24.6	236	1	ECSC_BACSU	PROTEIN ECSC.	7.57e+00
28	63	24.6	477	1	ANGT_MOUSE	ANGIOTENSINOGEN PRECUR	7.57e+00
29	63	24.6	639	1	TET9_ENTFA	TETRACYCLINE RESISTANC	7.57e+00
30	63	24.6	639	1	TET1_ENTFA	TETRACYCLINE RESISTANC	7.57e+00
31	63	24.6	639	1	TETW_STAAR	TETRACYCLINE RESISTANC	7.57e+00
32	62	24.2	150	1	EXBB_HAEDU	BIOPOLYMER TRANSPORT E	1.10e+01
33	62	24.2	612	1	UNC6_CAEEL	UNC-6 PROTEIN PRECURSO	1.10e+01
34	62	24.2	1464	1	NME1_RAT	GLUTAMATE [NMDA] RECEP	1.10e+01
35	62	24.2	1464	1	NME1_MOUSE	GLUTAMATE [NMDA] RECEP	1.10e+01
36	62	24.2	1466	1	SPA2_YEAST	SPA2 PROTEIN.	1.10e+01
37	62	24.2	1976	1	MYSO_HUMAN	MYOSIN HEAVY CHAIN, NO	1.10e+01
38	61	23.8	160	1	VG34_HSVEB	GENE 34 PROTEIN.	1.60e+01
39	61	23.8	270	1	URED_KLEPN	UREASE ACCESSORY PROTE	1.60e+01
40	61	23.8	270	1	URED_KLEAE	UREASE ACCESSORY PROTE	1.60e+01
41	61	23.8	406	1	Y129_HUMAN	HYPOTHETICAL PROTEIN K	1.60e+01
42	61	23.8	442	1	TBB_TRYBR	TUBULIN BETA CHAIN.	1.60e+01
43	61	23.8	483	1	PRPD_SALTY	PRPD PROTEIN.	1.60e+01
44	61	23.8	595	1	BETP_CORGL	GLYCINE BETAIN TRANSP	1.60e+01
45	61	23.8	1169	1	RAD5_YEAST	DNA REPAIR PROTEIN RAD	1.60e+01

## ALIGNMENTS

RESULT 1  
ID ICEB\_MOUSE STANDARD; PRT; 373 AA.  
AC P70343; O08735;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE CASPASE-11 PRECURSOR (EC 3.4.22.-) (ICH-3 PROTEASE).  
GN CASP11 OR ICH3 OR CASPL.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6 X CBA; TISSUE-THYMUS;  
RX MEDLINE; 96355393.  
RA WANG S., MIURA M., JUNG Y.-K., ZHU H., GAGLIARDINI V., SHI L.,  
RT GREENBERG A.H., YUAN J.;  
RT "Identification and characterization of Ich-3, a member of the  
interleukin-beta converting enzyme (ICE)/Ced-3 family and an  
upstream regulator of ICE.";  
RL J. BIOL. CHEM. 271:20580-20587(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C3H/AN;  
RX MEDLINE; 97190206.  
RA VAN DE CRAEN M., VANDENABEELE P., DECLERQ W., VAN DEN BRANDE I.,  
RA VAN LOO G., MOLEMAN F., SCHOTTE P., VAN CRIEKENGE W., BEYAERT R.,  
RA FIERIS W.;  
RT "Characterization of seven murine caspase family members.";  
RL FEBS LETT. 403:61-69(1997).  
CC -!- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES  
RESPONSIBLE FOR APOPTOSIS EXECUTION. PROMOTES IL-1 BETA PROCESSING  
BY ICE, SO MAY ALSO HAVE A ROLE IN INFLAMMATORY RESPONSES.  
CC -!- SUBUNIT: HETERODIMER OF A 20 KD (P20) AND A 10 KD (P10) SUBUNIT.  
CC -!- TISSUE SPECIFICITY: MOSTLY EXPRESSED IN LUNG AND SPLEEN; WEAKER IN  
HEART AND LUNG; LITTLE IN LIVER, SKELETAL MUSCLE, KIDNEY AND  
TESTIS. NOT FOUND IN THE BRAIN.  
CC -!- INDUCTION: ACTIVITY INCREASED 30-FOLD BY ENDOTOXINS (LPS).  
CC -!- PTM: THE TWO SUBUNITS ARE DERIVED FROM THE PRECURSOR SEQUENCE BY A  
AUTOCATALYTIC MECHANISM.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE  
CASPASE FAMILY.  
CC -!- CAUTION: THIS PROTEIN COULD BE THE MOUSE ORTHOLOG OF HUMAN  
CASPASE-4.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -

```
DE PUTATIVE ACID PHOSPHATASE C05C10.4 (EC 3.1.3.2).
GN C05C10.4
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA MATTHEWS P.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN (2)
RP REVISIONS.
RC STRAIN-BRISTOL N2;
RA JONES S.J.M.;
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O = AN
CC ALCOHOL + ORTHOPHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE HISTIDINE ACID PHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z48178; E1297507; -.
DR WORMPEP; C05C10.4; CE17370.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
DR PFAM; PF00328; acid_phosphat; 1.
KW HYPOTHETICAL PROTEIN; HYDROLASE.
FT ACT_SITE 35 35 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT_SITE 78 78 BY SIMILARITY.
FT ACT_SITE 314 314 PROTON DONOR (BY SIMILARITY).
FT DISULFID 381 387 BY SIMILARITY.
FT DOMAIN 55 60 POLY-GLY.
SQ SEQUENCE 413 AA; 46617 MW; BE3B5FCE CRC32;

Query Match 27.7%; Score 71; DB 1; Length 413;
Best Local Similarity 45.0%; Pred. No. 3.18e-01;
Matches 9; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

Db 152 PDQCAASQNCPCTRYDLLOQ 171
QY 4 PD-CRQKTCSCRLYELH 22

RESULT 5
ID R10A-TRYBR STANDARD; PRT; 214 AA.
AC P53028;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 60S RIBOSOMAL PROTEIN L10A.
GN RPL10A
OS TRYPA NOSOMA BRUCEI RHODESIENSE.
OC EUKARYOTA; EUKLENOZOA; KINETOPLASTIDA; TRYPA NOSOMATIDAE; TRYPA NOSOMA.
RN (1)
RP SEQUENCE FROM N.A.
RA EL-SAYED N.M., DONELSON J.E.;
RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: BELONGS TO THE LIP FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M88463; G459886; -.
DR EMBL; M29603; G567228; -.
DR PIR; A37199; A37199.
DR MGI; MGI:96945; MCS.
KW MITOCHONDRION; SELENIUM; SPERM; TESTIS; SPERMATOGENESIS.
FT BINDING 7 7 SELENIUM.
FT BINDING 17 17 SELENIUM.
FT BINDING 34 34 SELENIUM.
SQ SEQUENCE 197 AA; 21015 MW; 4E36990C CRC32;

Query Match 27.0%; Score 69; DB 1; Length 197;
Best Local Similarity 71.4%; Pred. No. 7.21e-01;
Matches 10; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

DR EMBL; U34620; G1123023; -.
DR PROSITE; PS01199; RIBOSOMAL_L1; 1.
DR PFAM; PF00687; Ribosomal_L1; 1.
KW RIBOSOMAL PROTEIN.
SQ SEQUENCE 214 AA; 24597 MW; EALB6765 CRC32;

Query Match 27.3%; Score 70; DB 1; Length 214;
Best Local Similarity 42.1%; Pred. No. 4.80e-01;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Db 52 LPHVCRPRMTVCLICDLVH 70
QY 3 LPDCCRQKTCSCRLYELH 21

RESULT 6
ID MCS_MOUSE STANDARD; PRT; 197 AA.
AC P13265;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE SPERM MITOCHONDRIAL CAPSULE SELENOPROTEIN (MCS).
GN MCSP OR MCS.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RX MEDLINE; 93039675.
RA KARIMPOUR I., CUTLER M., SHIH D., SMITH J., KLEENE K.C.;
RT "Sequence of the gene encoding the mitochondrial capsule
RT selenoprotein of mouse sperm: identification of three in-phase TGA
RT selenocysteine codons."
RT DNA CELL BIOL. 11:693-699(1992).
RN (2)
RP SEQUENCE OF 55-197 FROM N.A.
RC TISSUE-TESTIS;
RX MEDLINE; 90152148.
RA KLEENE K.C., SMITH J., BOZORGZADEH A., HARRIS M., HAHN L.,
RA KARIMPOUR I., GERSTEL J.;
RT "Sequence and developmental expression of the mRNA encoding the
RT seleno-protein of the sperm mitochondrial capsule in the mouse."
RL DEV. BIOL. 137:395-402(1990).
CC -!- FUNCTION: STRUCTURAL PROTEIN OF THE SPERM MITOCHONDRIAL CAPSULE.
CC IMPORTANT FOR THE MAINTENANCE AND STABILIZATION OF THE CRESCENT
CC STRUCTURE OF THE SPERM MITOCHONDRIA.
CC -!- SUBCELLULAR LOCATION: KERATINOUS MITOCHONDRIAL CAPSULE.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -!- DEVELOPMENTAL STAGE: LATE MEIOTIC AND EARLY HAPLOID CELLS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M88463; G459886; -.
DR EMBL; M29603; G567228; -.
DR PIR; A37199; A37199.
DR MGI; MGI:96945; MCS.
KW MITOCHONDRION; SELENIUM; SPERM; TESTIS; SPERMATOGENESIS.
FT BINDING 7 7 SELENIUM.
FT BINDING 17 17 SELENIUM.
FT BINDING 34 34 SELENIUM.
SQ SEQUENCE 197 AA; 21015 MW; 4E36990C CRC32;

Query Match 27.0%; Score 69; DB 1; Length 197;
Best Local Similarity 71.4%; Pred. No. 7.21e-01;
Matches 10; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
```



Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Db 215 SCSLYPVYHGS AKNIGIKQL 235  
QY 13 SCRLYELHAGNHAAGILT 33

RESULT 10  
ID ALC\_RABIT STANDARD; PRT; 299 AA.

AC P01879;  
DT 21-JUL-1986 (REL. 01, CREATED)  
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)  
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)  
DE IG ALPHA CHAIN C REGION (FRAGMENT).  
OS ORYCTOLAGUS CUNICULUS (RABBIT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 84144039.

RA KNIGHT K.L., MARTENS C.L., STOKLOS A.C.M., SCHNEIDERMAN R.D.;  
RT "Genes encoding alpha-heavy chains of rabbit IgA: characterization of  
RT cDNA encoding IGA-g subclass alpha-chains."  
RL NUCLEIC ACIDS RES. 12:1657-1670(1984).

CC -!- FUNCTION: IG ALPHA IS THE MAJOR IMMUNOGLOBULIN CLASS IN BODY  
CC SECRETIONS. IT MAY SERVE BOTH TO DEFEND AGAINST LOCAL INFECTION  
CC AND TO PREVENT ACCESS OF FOREIGN ANTIGENS TO THE GENERAL  
CC IMMUNOLOGIC SYSTEM.  
CC -!- THIS IMMUNOGLOBULIN BELONGS TO THE IGA-G SUBCLASS. IT WAS ISOLATED  
CC FROM A RABBIT HOMOZYGOUS FOR A2, N80, DE12,15, F71, G75 HEAVY  
CC CHAIN HAPLOTYPE.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).

CC -----  
CC EMBL; X00353; G1576; -.  
DR PIR; A02174; AHRB.  
DR PROSITE; PS00290; IG\_MHC; 2.  
DR PFAM; PF00047; ig; 2.  
KW IMMUNOGLOBULIN C REGION.  
FT NON\_TER 1  
SQ SEQUENCE 299 AA; 32256 MW; 594CED7C CRC32;

Query Match 26.2%; Score 67; DB 1; Length 299;  
Best Local Similarity 50.0%; Pred. No. 1.61e+00;  
Matches 7; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

Db 44 PPDCCPANSCCTC 57  
QY 2 PLPDCCROKTC-SC 14

RESULT 11  
ID REIL\_SCHPO STANDARD; PRT; 923 AA.

AC Q92380;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE MEIOTIC RECOMBINATION PROTEIN RECI1.  
GN RECI1.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;  
OC SCHIZOSACCHAROMYCETACEAE; SCHIZOSACCHAROMYCETACEAE;  
OC SCHIZOSACCHAROMYCETES.

RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97231330.  
RA LI Y.F., NUMATA M., WAHLS W.P., SMITH G.R.;

RT "Region-specific meiotic recombination in Schizosaccharomyces pombe:  
RT the rec11 gene.";  
RL MOL. MICROBIOL. 23:869-878(1997).

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).

CC -----  
CC EMBL; U70737; G1619901; -.  
DR MEIOSIS.  
KW SEQUENCE

SQ SEQUENCE 923 AA; 107418 MW; B51C7725 CRC32;

Query Match 26.2%; Score 67; DB 1; Length 923;  
Best Local Similarity 35.7%; Pred. No. 1.61e+00;  
Matches 10; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

Db 232 CCDIMRCLCLIVNKLSEKSNQTAELVL 259  
QY 6 CCRQKTCSCRLYELHAGNHAAGILT 33

RESULT 12  
ID GALE\_RAT STANDARD; PRT; 347 AA.

AC P18645;  
DT 01-NOV-1990 (REL. 16, CREATED)  
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE UDP-GLUCOSE 4-EPIPERASE (EC 5.1.3.2) (GALACTOWALDENASE) (UDP-  
DE GALACTOSE 4-EPIPERASE).  
GN GALE.  
OS RATIUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]  
RP SEQUENCE FROM N.A.

RC TISSUE-SKELETAL MUSCLE;  
RX MEDLINE; 90384840.  
RA ZESCHNIGK M., WILCKEN-BERGEMANN B., STARZINSKI-POWITZ A.;  
RT "cDNA from rat cells with reconstitutive galactose-epimerase activity  
RT in E. coli."  
RL NUCLEIC ACIDS RES. 18:5289-5289(1990).

CC -!- FUNCTION: CATALYZES TWO DISTINCT BUT ANALOGOUS REACTIONS: THE  
CC EPIMERIZATION OF UDP-GLUCOSE TO UDP-GALACTOSE AND THE  
CC EPIMERIZATION OF UDP-N-ACETYLGLUCOSAMINE TO UDP-N-  
CC ACETYLGLUCOSAMINE.  
CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE = UDP-GALACTOSE.

CC -!- COFACTOR: NAD.  
CC -!- PATHWAY: GALACTOSE METABOLISM.  
CC -!- SIMILARITY: WITH OTHER GALACTOWALDENASES FROM EUKARYOTIC AND  
CC PROKARYOTIC ORIGIN.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).

CC -----  
CC EMBL; X53949; G57792; -.  
DR PIR; S11223; S11223.  
DR PFAM; PF00106; adh\_short; 1.  
DR HSSP; P09147; 1KVS.  
KW ISOMERASE; NAD; GALACTOSE METABOLISM.  
FT NP\_BIND 4 35 NAD (POTENTIAL).  
SQ SEQUENCE 347 AA; 38225 MW; 3EC2E611 CRC32;

Query Match 25.8%; Score 66; DB 1; Length 347;  
Best Local Similarity 53.8%; Pred. No. 2.38e+00;

Search completed: Fri Aug 20 21:07:07 1999  
Job time : 19 secs.

\*\*\*\*\*

W P E R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Fri Aug 20 21:06:09 1999; MasPar time 5.02 Seconds  
Tabular output not generated. 263.383 Million cell updates/sec

Title: >US-08-938-548B-8  
Description: (1-33) from US08938548B.pep  
Perfect Score: 256  
Sequence: 1 OPLPDCRCROKTCSCRYELLHGAGNHAGILTL 33  
Scoring table: PAM 150  
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 30.710; Variance 49.199; scale 0.624

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	85	33.2	915	2	T00728	hypothetical protein	3.46e-03
2	72	28.1	147	2	S24303	hypothetical protein	6.20e-01
3	70	27.3	870	2	A41130	dystrophin homolog	1.32e+00
4	69	27.0	93	2	S60079	oct2 protein isoform	1.91e+00
5	69	27.0	143	2	A37199	sperm mitochondrial c	1.91e+00
6	69	27.0	533	2	T01864	hypothetical protein	1.91e+00
7	68	26.6	641	2	JN0800	tetracycline-minocycl	2.76e+00
8	67	26.2	299	1	ARRB	Ig alpha chain C regi	3.98e+00
9	67	26.2	338	2	S09276	Ig alpha chain C regi	3.98e+00
10	66	25.8	347	2	S11223	UDPGlucose 4-epimeras	5.71e+00
11	66	25.8	425	2	I56329	gene D3 protein - mou	5.71e+00
12	66	25.8	479	2	B69764	transcription regulat	5.71e+00
13	65	25.4	155	2	C64314	conserved hypothetical	8.17e+00
14	65	25.4	197	2	I46413	keratin KAP5.5 - sh	8.17e+00
15	65	25.4	589	2	S33920	beta-fructofuranosida	8.17e+00
16	65	25.4	806	2	T01164	hypothetical protein	8.17e+00
17	64	25.0	48	1	A44664	omega-agatoxin IIV	1.16e+01
18	64	25.0	76	2	A54252	omega-agatoxin III	1.16e+01
19	64	25.0	76	2	B54252	omega-agatoxin III	1.16e+01
20	64	25.0	76	2	A24335	omega-agatoxin IIIA	1.16e+01
21	64	25.0	103	2	B64052	cell division protein	1.16e+01
22	64	25.0	131	2	A46298	pigment deposition co	1.16e+01
23	64	25.0	132	2	I37143	agouti protein precur	1.16e+01

24	64	25.0	369	2	B64835	probable iron-sulfur-	1.16e+01
25	64	25.0	493	2	S49175	cysteine proteinase (	1.16e+01
26	63	24.6	236	2	H69619	hypothetical protein	1.65e+01
27	63	24.6	360	2	S09271	Ig alpha chain C regi	1.65e+01
28	63	24.6	396	2	B57479	amino acid transport	1.65e+01
29	63	24.6	477	2	A29978	angiotensin precursor	1.65e+01
30	63	24.6	577	2	T01052	heatshock protein dna	1.65e+01
31	63	24.6	639	2	A60833	tetracycline resistan	1.65e+01
32	63	24.6	639	2	A56779	tetracycline resistan	1.65e+01
33	63	24.6	639	2	S13142	tetracycline resistan	1.65e+01
34	62	24.2	138	2	G70431	hypothetical protein	2.33e+01
35	62	24.2	157	2	JC5551	hypothetical 17.9K pr	2.33e+01
36	62	24.2	612	2	JH0799	laminin-related prote	2.33e+01
37	62	24.2	722	2	B61231	myosin heavy chain, n	2.33e+01
38	62	24.2	1464	2	A43274	N-methyl D-aspartate	2.33e+01
39	62	24.2	1464	2	S47555	N-methyl-D-aspartate	2.33e+01
40	62	24.2	1464	2	S29159	glutamate receptor, N	2.33e+01
41	62	24.2	1466	2	A36426	SPA2 protein - yeast	2.33e+01
42	61	23.8	160	1	WZBEC7	gene 34 protein - equ	3.28e+01
43	61	23.8	442	1	UBUTB	tubulin beta chain -	3.28e+01
44	61	23.8	1169	1	S64859	DNA repair protein RA	3.28e+01
45	61	23.8	2007	1	B43402	myosin heavy chain-B,	3.28e+01

#### ALIGNMENTS

RESULT 1  
ENTRY T00728 #type complete  
TITLE hypothetical protein F22013.23 - Arabidopsis thaliana  
ORGANISM #formal\_name Arabidopsis thaliana #common\_name mouse-ear cress  
DATE 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 12-Feb-1999  
ACCESSIONS T00728  
REFERENCE Z14200  
#authors Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; Conway, A.; Kurtz, D.; Oji, O.; Shen, Y.K.; Toriumi, M.; Vysotskaia, V.; Yu, G.; Davis, R.W.; Federspiel, N.A.; Theologis, A.; Ecker, J.R.  
#submission submitted to the EMBL Data Library, April 1998  
#description Genomic sequence for Arabidopsis thaliana BAC F22013.  
#accession T00728  
#status preliminary; translated from GB/EMBL/DBJ  
#molecule\_type DNA  
#residues 1-915 #label SHI  
#cross-references EMBL:AC003981; NID:g3063438; PID:g3063461  
GENETICS  
#map\_position 1  
#introns 222/3; 326/1; 619/3; 889/3  
SUMMARY #length 915 #molecular-weight 103046 #checksum 9587  
Query Match 33.2%; Score 85; DB 2; Length 915;  
Best Local Similarity 61.5%; Pred. No. 3.46e-03;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
Db 794 ERTCSRCRFDLLH 806  
QY :||||| :|||  
9 QKTCSCRYELLH 21

RESULT 2  
ENTRY S24303 #type complete  
TITLE hypothetical protein H19-3' - mouse  
ORGANISM #formal\_name Mus musculus #common\_name house mouse  
DATE 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 10-Sep-1997  
ACCESSIONS S24303  
REFERENCE S24302  
#authors Poirier, F.; Chan, C.T.J.; Timmons, P.M.; Robertson, E.J.; Evans, M.J.; Rigby, P.W.J.  
#journal Development (1991) 113:1105-1114  
#title The murine H19 gene is activated during embryonic stem cell differentiation in vitro and at the time of implantation in

```

7
RESULT 7
ENTRY 7
TITLE tetacycline-minocycline resistance protein - Listeria
ORGANISM #formal_name Listeria monocytogenes
DATE 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change
28-Aug-1998
ACCESSIONS JN0800
REFERENCE JN0800
#authors Charpentier, E.; Gerbaud, G.; Courvalin, P.
#journal Gene (1993) 131:27-34
#title Characterization of a new class of tetracycline-resistance
#cross-references MUID:93380670
#accession JN0800
##molecule_type DNA
##residues 1-641 ##label CHA
##cross-references GB:L09756; NID:9406084; PID:9406085
##experimental_source strain BM4210
##note the start codon is "TTC" encoding a "Leu" amino acid in
the paper

GENETICS
#gene tet(S)
#genome plasmid
#start_codon TTG
CLASSIFICATION #superfamily translation elongation factor G; translation
KEYWORDS elongation factor Tu homology
FEATURE antibiotic resistance; GTP binding; P-loop
4-131
10-17 #domain translation elongation factor Tu homology #label
EFL\
128-131 #region nucleotide-binding motif A (P-loop)\
220-222 #region GTP-binding NKXD motif\
16,17,55,128,129, #region GTP-binding SAK/L motif\
131,220 #binding_site Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser)
#status predicted
SUMMARY #length 641 #molecular_weight 73013 #checksum 8738
Query Match 26.6%; Score 68; DB 2; Length 641;
Best Local Similarity 42.9%; Pred. No. 2.76e+00;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Db 210 SCSLYPVVHGSAXKNIGIKQL 230
|||:||||:|||||
Qy 13 SCLYELLHGAGNHAAGILT 33

8
RESULT 8
ENTRY 8
TITLE AHRB #type fragment
ORGANISM Ig alpha chain C region - rabbit (fragment)
#formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change
20-Mar-1998
ACCESSIONS A02174
REFERENCE A02174
#authors Knight, K.L.; Martens, C.L.; Stoklosa, C.M.; Schneiderman,
R.D.
#journal Nucleic Acids Res. (1984) 12:1657-1670
#title Genes encoding alpha-heavy chains of rabbit IgA:
characterization of cDNA encoding IgA-g subclass
alpha-chains.
#cross-references MUID:84144059
#accession A02174
##molecule_type mRNA
##residues 1-299 ##label KNI
##cross-references GB:X00353; NID:gl575; PID:gl576
COMMENT This immunoglobulin belongs to the IgA-g subclass. It was isolated
from a rabbit homozygous for a2, n80, de12,15, f71, g75 heavy
chain haplotype.

COMPLEX
An immunoglobulin heterotetramer subunit consists of two
identical light (kappa or lambda) and two identical heavy
(alpha, delta, epsilon, gamma, or mu) chains usually
stabilized by interchain disulfide bonds. In some cases,
such as IgA and IgM, the subunits associate into larger
oligomers.
#superfamily immunoglobulin C region; immunoglobulin homology
duplication; glycoprotein; heterotetramer; immunoglobulin;
plasma

CLASSIFICATION
KEYWORDS #domain immunoglobulin homology #label IGG1\
189-261 #domain immunoglobulin homology #label IGG2\
38,286 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 299 #checksum 2361
Query Match 26.2%; Score 67; DB 1; Length 299;
Best Local Similarity 50.0%; Pred. No. 3.98e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

Db 44 PFPDCCPANSCCTC 57
|:||||:|:|:|
Qy 2 PLPDCCCRQKTC-SC 14

RESULT 9
ENTRY 9
TITLE S09276 #type fragment
ORGANISM Ig alpha chain C region - rabbit (fragment)
#formal_name Oryctolagus cuniculus #common_name domestic
rabbit
DATE 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change
16-Aug-1996
ACCESSIONS S09276
REFERENCE S09264
#authors Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.
#journal EMBO J. (1989) 8:4041-4047
#title The IgA heavy-chain gene family in rabbit: cloning and
sequence analysis of 13 C-alpha genes.
#cross-references MUID:90076124
#accession S09276
##status not compared with conceptual translation
##molecule_type DNA
##residues 1-338 ##label BUR
CLASSIFICATION #superfamily immunoglobulin C region; immunoglobulin homology
KEYWORDS immunoglobulin
FEATURE #domain immunoglobulin homology #label IGG2
228-300
SUMMARY #length 338 #checksum 2169
Query Match 26.2%; Score 67; DB 2; Length 338;
Best Local Similarity 50.0%; Pred. No. 3.98e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

Db 83 PFPDCCPANSCCTC 96
|:||||:|:|:|
Qy 2 PLPDCCCRQKTC-SC 14

RESULT 10
ENTRY 10
TITLE S11223 #type complete
ORGANISM UDPglucose 4-epimerase (EC 5.1.3.2) - rat
ALTERNATE_NAMES UDPgalactose 4-epimerase
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change
08-Sep-1997
ACCESSIONS S11223
REFERENCE S11223
#authors Zeschnick, M.; von Wilcken-Bergmann, B.; Starzinski-Powitz,
A.
#journal Nucleic Acids Res. (1990) 18:5289
#title cDNA from rat cells with reconstitutive galactose-epimerase
activity in E. coli.
#cross-references MUID:90384840

```

```

#cross-references MUID:96337999
#accession C64314
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA
#residues 1-155 #label BUL
#cross-references GB:U67469; GB:L77117; NID:G1590890; PID:G1498882;
TIGR:MJ0115

GENETICS
#map_position REV113249-112782
CLASSIFICATION #superfamily Methanococcus jannaschii conserved hypothetical
protein Mj0115
SUMMARY #length 155 #molecular-weight 17727 #checksum 834

Query Match 25.4%; Score 65; DB 2; Length 155;
Best Local Similarity 53.3%; Pred. NO. 8.17e+00;
Matches 8; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

Db 72 CCKIKPCPYRDYEL 86
||: ||: |||
QY 6 CCK-OKTCSRLYEL 19

RESULT 14
ENTRY I46413 #type fragment
TITLE keratin KAP5.5 - sheep (fragment)
ORGANISM #formal_name Ovis orientalis aries, Ovis ammon aries
#common_name domestic sheep
DATE 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change
13-Sep-1998
ACCESSIONS I46413; S34216
REFERENCE I46412
#authors Jenkins, B.J.; Powell, B.C.
#journal J. Invest. Dermatol. (1994) 103:310-317
#title Differential expression of genes encoding a cysteine-rich
keratin family in the hair cuticle.
#cross-references MUID:94358466
#accession I46413
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-197 #label JEN
#cross-references EMBL:X73435; NID:G313721; PID:G313722

GENETICS
#gene KRTAP5.5
CLASSIFICATION #superfamily ultra-high-sulfur keratin
SUMMARY #length 197 #checksum 2787

Query Match 25.4%; Score 65; DB 2; Length 197;
Best Local Similarity 58.3%; Pred. NO. 8.17e+00;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 87 VPVCCFVPACSC 98
: ||| :|||
QY 3 LPDCCRQKTCSC 14

RESULT 15
ENTRY S33920 #type complete
TITLE beta-fructofuranosidase (EC 3.2.1.26) precursor - Aspergillus
niger
ALTERNATE_NAMES beta-D-fructofuranosidase; invertase; saccharase
ORGANISM #formal_name Aspergillus niger
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
13-Sep-1998
ACCESSIONS S33920; S36775
REFERENCE S33920
#authors Boddy, L.M.; Berge, T.; Barreau, C.; Vainstein, M.H.;
Dobson, M.J.; Ballance, D.J.; Peberdy, J.F.
#journal Curr. Genet. (1993) 24:60-66
#title Purification and characterisation of an Aspergillus niger
invertase and its DNA sequence.
#cross-references MUID:93365038

```

```

#accession S33920
#molecule_type DNA
#residues 1-589 #label BOD1
#cross-references EMBL:L06844
#accession S36775
#molecule_type protein
#residues 24-35;79-92;369-383;402-413 #label BOD2

GENETICS
#gene suc1
CLASSIFICATION #superfamily beta-fructofuranosidase
KEYWORDS glycoprotein; glycosidase; hydrolase; phosphoprotein
FEATURE
1-23 #domain signal sequence #status predicted #label SIG\
24-589 #product beta-fructofuranosidase #status experimental
#label MAT\

36,42,170,188,211,
254,259,318,322,
388,463,518,527
#binding_site carbohydrate (Asn) (covalent) #status
predicted\
70,92 #binding_site phosphate (Thr) (covalent) #status
predicted\
458,475,490 #binding_site phosphate (Ser) (covalent) #status
predicted

SUMMARY #length 589 #molecular-weight 63650 #checksum 1032

Query Match 25.4%; Score 65; DB 2; Length 589;
Best Local Similarity 42.1%; Pred. NO. 8.17e+00;
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

Db 548 RLFDVLNG-GEQAETLDL 565
||::||: ||: ||
QY 15 RUYELHAGNHAAGILT 33

```

Search completed: Fri Aug 20 21:06:31 1999  
Job time : 22 secs.

\*\*\*\*\*  
W P S R L H  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:08:23 1999; Maspar time 1.89 Seconds  
Tabular output not generated.

Title: >US-08-938-548B-8  
Description: (1-33) from US08938548B.peg  
Perfect Score: 256  
Sequence: 1 QPLPDCRQKTCSCRLYELLHGAGNHAAGILTL 333

Scoring table: PAM 150  
Gap 11

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 21.693; Variance 81.714; scale 0.265

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	66	25.8	39	3	PCT-US95-0	Sequence 40, Applicati	3.62e+01
2	66	25.8	39	1	US-08-249-	Sequence 40, Applicati	3.62e+01
3	66	25.8	39	2	US-08-734-	Sequence 40, Applicati	3.62e+01
4	66	25.8	39	2	US-08-469-	Sequence 40, Applicati	3.62e+01
5	66	25.8	39	1	US-08-036-	Sequence 40, Applicati	3.62e+01
6	66	25.8	39	3	PCT-US94-0	Sequence 40, Applicati	3.62e+01
7	66	25.8	39	1	US-08-469-	Sequence 40, Applicati	3.62e+01
8	66	25.8	39	2	US-08-469-	Sequence 40, Applicati	3.62e+01
9	66	25.8	348	3	PCT-US95-0	Sequence 2, Applicatio	3.62e+01
10	65	25.4	88	2	US-08-469-	Sequence 20, Applicati	4.41e+01
11	64	25.0	1167	1	US-08-485-	Sequence 6, Applicatio	5.37e+01
12	64	25.0	1167	2	US-08-590-	Sequence 6, Applicatio	5.37e+01
13	64	25.0	1168	1	US-08-620-	Sequence 9, Applicatio	5.37e+01
14	62	24.2	1464	2	US-08-231-	Sequence 11, Applicati	7.93e+01
15	62	24.2	1464	1	US-08-026-	Sequence 1, Applicatio	7.93e+01
16	61	23.8	42	1	US-08-137-	Sequence 23, Applicati	9.63e+01
17	61	23.8	42	1	US-08-487-	Sequence 23, Applicati	9.63e+01
18	61	23.8	42	1	US-08-477-	Sequence 23, Applicati	9.63e+01
19	61	23.8	42	1	US-08-480-	Sequence 23, Applicati	9.63e+01
20	61	23.8	59	1	US-08-233-	Sequence 51, Applicati	9.63e+01
21	60	23.4	222	4	5232425-6	Patent No. 5232425.	1.17e+02
22	60	23.4	334	1	US-08-279-	Sequence 10, Applicati	1.17e+02
23	59	23.0	283	1	US-08-658-	Sequence 2, Applicatio	1.41e+02

24	59	23.0	1019	1	US-08-296-	Sequence 4, Applicatio	1.41e+02
25	59	23.0	1019	2	US-08-596-	Sequence 4, Applicatio	1.41e+02
26	59	23.0	1083	1	US-08-296-	Sequence 2, Applicatio	1.41e+02
27	59	23.0	1083	2	US-08-596-	Sequence 2, Applicatio	1.41e+02
28	58	22.7	54	2	US-08-757-	Sequence 8, Applicatio	1.71e+02
29	58	22.7	131	2	US-08-757-	Sequence 10, Applicati	1.71e+02
30	58	22.7	132	2	US-08-757-	Sequence 11, Applicati	1.71e+02
31	58	22.7	132	2	US-08-757-	Sequence 7, Applicatio	1.71e+02
32	57	22.3	21	1	US-08-019-	Sequence 5, Applicatio	2.07e+02
33	57	22.3	36	2	US-08-451-	Sequence 8, Applicatio	2.07e+02
34	57	22.3	48	1	US-08-379-	Sequence 1, Applicatio	2.07e+02
35	57	22.3	227	2	US-08-698-	Sequence 1, Applicatio	2.07e+02
36	57	22.3	424	2	US-08-951-	Sequence 9, Applicatio	2.07e+02
37	57	22.3	431	2	US-08-928-	Sequence 2, Applicatio	2.07e+02
38	57	22.3	724	3	PCT-US94-1	Sequence 62, Applicati	2.07e+02
39	57	22.3	790	1	US-08-363-	Sequence 2, Applicatio	2.07e+02
40	57	22.3	1184	1	US-08-446-	Sequence 20, Applicati	2.07e+02
41	57	22.3	1184	1	US-08-446-	Sequence 20, Applicati	2.07e+02
42	57	22.3	1184	2	US-08-805-	Sequence 20, Applicati	2.07e+02
43	57	22.3	1187	3	PCT-US95-1	Sequence 8, Applicatio	2.07e+02
44	57	22.3	1187	1	US-08-097-	Sequence 13, Applicati	2.07e+02
45	57	22.3	1187	1	US-08-357-	Sequence 8, Applicatio	2.07e+02

ALIGNMENTS

RESULT 1  
ID PCT-US95-06846A-40 STANDARD; PRT; 39 AA.  
AC XXXXXX  
XX  
DT  
XX  
DE  
XX  
XX  
Sequence 40, Application PC/TUS9506846A  
CC  
Sequence 40, Application PC/TUS9506846A  
CC  
GENERAL INFORMATION:  
CC APPLICANT: Goodearl, Andrew David; Stroobant, Paul;  
CC APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchionni, Mark;  
CC APPLICANT: Chen, Maio Su; Hiles, Ian  
CC TITLE OF INVENTION: Glial Mitogenic Factors, Their  
CC TITLE OF INVENTION: Preparation and Use  
CC NUMBER OF SEQUENCES: 178  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Felfe & Lynch  
CC STREET: 805 Third Avenue  
CC CITY: New York City  
CC STATE: New York  
CC COUNTRY: USA  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
CC COMPUTER: IBM  
CC OPERATING SYSTEM: PC-DOS  
CC SOFTWARE: Wordperfect  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US95/06846A  
CC FILING DATE: 25-MAY-1995  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/036,555  
CC FILING DATE: 24-MAR-1993  
CC FILING DATE: 26-MAY-1994  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/965,173  
CC FILING DATE: 23-OCT-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/940,389  
CC FILING DATE: 03-SEP-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/907,138

CC APPLICATION NUMBER: 08/470,335  
CC FILING DATE: 06-JUN-1995  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/036,555  
CC FILING DATE: 03-MAR-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/965,173  
CC FILING DATE: 23-OCT-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/940,389  
CC FILING DATE: 03-SEP-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/907,138  
CC FILING DATE: 30-JUN-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/863,703  
CC FILING DATE: 03-APR-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: UK 91 07566.3  
CC FILING DATE: 10-APR-1991  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Bieker-Brady, Kristina  
CC REGISTRATION NUMBER: 39,109  
CC REFERENCE/DOCKET NUMBER: 04585/00200P  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 428-0200  
CC TELEFAX: (617) 428-7045  
CC TELEX:  
CC INFORMATION FOR SEQ ID NO: 40:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 39  
CC TYPE: amino acid  
CC STRANDEDNESS:  
CC TOPOLOGY: linear  
CC SEQUENCE 39 AA; 3760 MW; 6521 CN;

Query Match 25.8%; Score 66; DB 2; Length 39;  
Best Local Similarity 70.0%; Pred. No. 3.62e+01;  
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Db 8 CCR-TTCACR 16  
||| ||:||  
Qy 6 CCRQKTCSCR 15

RESULT 4  
ID US-08-469-660-40 STANDARD; PRT; 39 AA.

XX AC xxxxxx

XX DF

XX DT

XX DE

Sequence 40, Application US/08469660

XX Sequence 40, Application US/08469660

CC Patent No. 5876973

CC GENERAL INFORMATION:

CC APPLICANT: Gwynne, David I.; Marchionni, Mark;

CC APPLICANT: McBurney, Robert N.

CC TITLE OF INVENTION: INHIBITORS OF CELL PROLIFERATION,

CC TITLE OF INVENTION: THEIR PREPARATION AND USE

CC NUMBER OF SEQUENCES: 184

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Fish & Richardson

CC STREET: 225 Franklin Street

CC CITY: Boston

CC STATE: Massachusetts

CC ZIP: 02111-2804

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

CC COMPUTER: IBM

CC OPERATING SYSTEM: PC-DOS

CC SOFTWARE: Wordperfect

CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/469,660  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/011,396  
CC FILING DATE: 29-JAN-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/984,085  
CC FILING DATE: 01-DEC-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/951,747  
CC FILING DATE: 25-SEP-1992  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 07/927,337  
CC FILING DATE: 10-AUG-1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Clark, Paul T.  
CC REGISTRATION NUMBER: 30,162  
CC REFERENCE/DOCKET NUMBER: 04585/017004  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (617) 542-5070  
CC TELEFAX: 200154  
CC INFORMATION FOR SEQ ID NO: 40:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 39  
CC TYPE: amino acid  
CC STRANDEDNESS:  
CC TOPOLOGY: linear  
CC SEQUENCE 39 AA; 3760 MW; 6521 CN;

Query Match 25.8%; Score 66; DB 2; Length 39;  
Best Local Similarity 70.0%; Pred. No. 3.62e+01;  
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Db 8 CCR-TTCACR 16  
||| ||:||  
Qy 6 CCRQKTCSCR 15

RESULT 5  
ID US-08-036-555B-40 STANDARD; PRT; 39 AA.

XX AC xxxxxx

XX DF

XX DT

XX DE

Sequence 40, Application US/08036555B

XX Sequence 40, Application US/08036555B

CC Patent No. 5530109

CC GENERAL INFORMATION:

CC APPLICANT: Goodearl, Andrew; Stroobant, Paul;

CC APPLICANT: Minghetti, Luisa; Waterfield, Michael; Marchioni, Mark;

CC APPLICANT: Chen, Maio Su; Hiles, Ian

CC TITLE OF INVENTION: Glial Mitogenic Factors, Their

CC TITLE OF INVENTION: Preparation and Use

CC NUMBER OF SEQUENCES: 184

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Felfe & Lynch

CC STREET: 805 Third Avenue

CC CITY: New York City

CC STATE: New York

CC COUNTRY: USA

CC ZIP: 10022

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

CC COMPUTER: IBM

CC OPERATING SYSTEM: PC-DOS

CC SOFTWARE: Wordperfect

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/036,555B

CC FILING DATE: 24-MAR-1993

CC APPLICATION NUMBER: 07/907,138  
 CC FILING DATE: 30-JUN-1992  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 07/863,703  
 CC FILING DATE: 03-APRIL-1992  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: U.K. 91 07566.3  
 CC FILING DATE: 10-APRIL-1991  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Tsal, Christine H.  
 CC REGISTRATION NUMBER: 34,266  
 CC REFERENCE/DOCKET NUMBER: LUD 5250.4  
 CC TELEPHONE: (212) 688-5200  
 CC TELEFAX: (212) 838-3884  
 CC INFORMATION FOR SEQ ID NO: 40:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 39  
 CC TYPE: amino acid  
 CC STRANDEDNESS:  
 CC TOPOLOGY: linear  
 CC SEQUENCE 39 AA: 3760 MW; 6521 CN;  
 CC  
 CC Query Match 25.8%; Score 66; DB 1; Length 39;  
 CC Best Local Similarity 70.0%; Pred. No. 3.62e+01;  
 CC Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 1;  
 Db 8 CCR-TTCACR 16  
 QY 6 CCROKTCSCR 15  
 RESULT 8  
 ID US-08-469-526A-40 STANDARD; PRT; 39 AA.  
 XX  
 AC xxxxxx  
 DT  
 XX  
 DE  
 XX  
 Sequence 40, Application US/08469526A  
 Sequence 40, Application US/08469526A  
 Patent No. 5792849  
 GENERAL INFORMATION:  
 CC APPLICANT: Goodearl, Andrew  
 CC APPLICANT: Stroobant, Paul  
 CC APPLICANT: Minghetti, Luisa  
 CC APPLICANT: Waterfield, Michael  
 CC APPLICANT: Marchionni, Mark  
 CC APPLICANT: Chen, Maio Su  
 CC APPLICANT: Hiles, Ian  
 CC TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR  
 CC PREPARATION AND USE  
 CC NUMBER OF SEQUENCES: 187  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: Clark & Elbing LLP  
 CC STREET: 176 Federal Street  
 CC CITY: Boston  
 CC STATE: MA  
 CC COUNTRY: USA  
 CC ZIP: 02110  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: Diskette  
 CC COMPUTER: IBM Compatible  
 CC OPERATING SYSTEM: DOS  
 CC SOFTWARE: FastSeq for Windows Version 2.0  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: US/08/469,526A  
 CC FILING DATE: 06 June 1995  
 CC CLASSIFICATION: 435  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER: 08/036,555  
 CC FILING DATE: 24-MAR-1993

CC APPLICATION NUMBER: 07/965,173  
 CC FILING DATE: 23-OCT-1992  
 CC PRIOR APPLICATION NUMBER: 07/940,389  
 CC FILING DATE: 03-SEP-1992  
 CC APPLICATION NUMBER: 07/907,138  
 CC FILING DATE: 03-JUN-1992  
 CC APPLICATION NUMBER: 07/863,703  
 CC FILING DATE: 03-APRIL-1992  
 CC APPLICATION NUMBER: U.K. 91 07566.3  
 CC FILING DATE: 10-APR-1991  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: Bicker-Brady, Kristina  
 CC REGISTRATION NUMBER: 39,109  
 CC REFERENCE/DOCKET NUMBER: 04585/00200A  
 CC TELEPHONE: 617-428-0200  
 CC TELEFAX: 617-428-7045  
 CC INFORMATION FOR SEQ ID NO: 40:  
 CC SEQUENCE CHARACTERISTICS:  
 CC LENGTH: 39  
 CC TYPE: amino acid  
 CC STRANDEDNESS: single  
 CC TOPOLOGY: linear  
 CC SEQUENCE 39 AA: 3760 MW; 6521 CN;  
 CC  
 CC Query Match 25.8%; Score 66; DB 2; Length 39;  
 CC Best Local Similarity 70.0%; Pred. No. 3.62e+01;  
 CC Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 1;  
 Db 8 CCR-TTCACR 16  
 QY 6 CCROKTCSCR 15  
 RESULT 9  
 ID PCT-US95-05785-2 STANDARD; PRT; 348 AA.  
 XX  
 AC xxxxxx  
 DT  
 XX  
 DE  
 XX  
 Sequence 2, Application PC/TUS9505785  
 Sequence 2, Application PC/TUS9505785  
 GENERAL INFORMATION:  
 CC APPLICANT: JI, H, ET AL.  
 CC TITLE OF INVENTION: HUMAN UDP GALACTOSE-4-EPIMERASE  
 CC NUMBER OF SEQUENCES: 10  
 CC CORRESPONDENCE ADDRESS:  
 CC ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
 CC ADDRESSEE: CECCHI, STEWART & OLSTEIN  
 CC STREET: 6 BECKER FARM ROAD  
 CC CITY: ROSELAND  
 CC STATE: NEW JERSEY  
 CC COUNTRY: USA  
 CC ZIP: 07068  
 CC COMPUTER READABLE FORM:  
 CC MEDIUM TYPE: 3.5 INCH DISKETTE  
 CC COMPUTER: IBM PS/2  
 CC OPERATING SYSTEM: MS-DOS  
 CC SOFTWARE: WORD PERFECT 5.1  
 CC CURRENT APPLICATION DATA:  
 CC APPLICATION NUMBER: PCT/US95/05785  
 CC FILING DATE: Concurrently  
 CC CLASSIFICATION:  
 CC PRIOR APPLICATION DATA:  
 CC APPLICATION NUMBER:  
 CC FILING DATE:  
 CC ATTORNEY/AGENT INFORMATION:  
 CC NAME: FERRARO, GREGORY D.  
 CC REGISTRATION NUMBER: 36,134  
 CC REFERENCE/DOCKET NUMBER: 325800-305  
 CC TELECOMMUNICATION INFORMATION:



CC TELEPHONE: (904) 375-8100  
CC TELEFAX: (904) 372-5800  
CC INFORMATION FOR SEQ ID NO: 6:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1167 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 1167 AA; 131657 MW; 7095219 CN;

Query Match 25.0%; Score 64; DB 1; Length 1167;  
Best Local Similarity 45.5%; Pred. No. 5.37e-01;  
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 873 PNCCKPAAQC 883  
QY 4 PDCCKKTCSC 14  
|:|:|:|:|

RESULT 12  
ID US-08-590-554A-6 STANDARD; PRT; 1167 AA.  
XX AC xxxxxx

Sequence 6, Application US/08590554A  
Sequence 6, Application US/08590554A  
Patent No. 5831011  
GENERAL INFORMATION:  
CC APPLICANT: Payne, Jewel  
CC APPLICANT: Narva, Kenneth E.  
CC APPLICANT: Fu, Jenny M.  
CC TITLE OF INVENTION: NO. 5831011el Bacillus thuringiensis Genes Encoding  
CC TITLE OF INVENTION: Nematode-Active Toxins  
CC NUMBER OF SEQUENCES: 6  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: David R. Saliwanchik  
CC STREET: 2421 N.W. 41st Street, Suite A-1  
CC CITY: Gainesville  
CC STATE: Florida  
CC COUNTRY: USA  
CC ZIP: 32606  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent in Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/590.554A  
CC FILING DATE: 21-MAR-1996  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/485,568  
CC FILING DATE: 7-JUN-1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/310,197  
CC FILING DATE: 21-SEP-1994  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/918,345  
CC FILING DATE: 21-JUL-1992  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/558,738  
CC FILING DATE: 27-JUL-1990  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/357,698  
CC FILING DATE: 16-DEC-1994

CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/176,403  
CC FILING DATE: 30-DEC-1993  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/999,053  
CC FILING DATE: 14-JUL-1993  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Saliwanchik, David R.  
CC REGISTRATION NUMBER: 31,794  
CC REFERENCE/DOCKET NUMBER: MA48DD2.C2  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (352) 375-8100  
CC TELEFAX: (352) 372-5800  
CC INFORMATION FOR SEQ ID NO: 6:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 1167 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 1167 AA; 131657 MW; 7095219 CN;

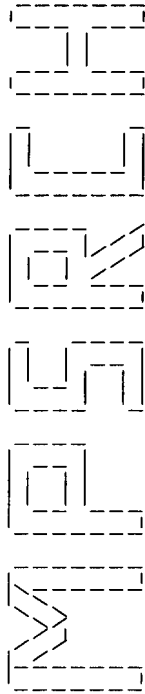
Query Match 25.0%; Score 64; DB 2; Length 1167;  
Best Local Similarity 45.5%; Pred. No. 5.37e+01;  
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 873 PNCCKPAAQC 883  
QY 4 PDCCKKTCSC 14  
|:|:|:|:|

RESULT 13  
ID US-08-620-717A-9 STANDARD; PRT; 1168 AA.  
XX AC xxxxxx  
XX DT  
XX DE  
XX SEQUENCE 9, Application US/08620717A  
XX Sequence 9, Application US/08620717A  
XX Patent No. 5670365  
XX GENERAL INFORMATION:  
XX APPLICANT: Feitelson, Jerald S.  
XX TITLE OF INVENTION: Identification of, and Uses For, Nematicidal  
XX TITLE OF INVENTION: Bacillus thuringiensis Genes, Toxins, and Isolates  
XX NUMBER OF SEQUENCES: 9  
XX CORRESPONDENCE ADDRESS:  
XX ADDRESSEE: Saliwanchik & Saliwanchik  
XX STREET: 2421 N.W. 41st Street, Suite A-1  
XX CITY: Gainesville  
XX STATE: Florida  
XX COUNTRY: USA  
XX ZIP: 32606  
XX COMPUTER READABLE FORM:  
XX MEDIUM TYPE: Floppy disk  
XX COMPUTER: IBM PC compatible  
XX OPERATING SYSTEM: PC-DOS/MS-DOS  
XX SOFTWARE: Patent in  
XX CURRENT APPLICATION DATA:  
XX APPLICATION NUMBER: US 08/620,717A  
XX FILING DATE:  
XX CLASSIFICATION: 435  
XX PRIOR APPLICATION DATA:  
XX APPLICATION NUMBER: US 08/540,104  
XX FILING DATE: 06-OCT-1995  
XX CLASSIFICATION: 435  
XX ATTORNEY/AGENT INFORMATION:  
XX NAME: Saliwanchik, David R.  
XX REGISTRATION NUMBER: 31,794



\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:05:04 1999; MasPar time 5.22 Seconds  
134.533 Million cell updates/sec

Tabular output not generated.

Title: >US-08-938-548B-8  
Description: (1-33) from US08938548B.pep  
Perfect Score: 256  
Sequence: 1 QPLPDCRCQKTCRSLYELHAGNHAAGILT 33  
Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 23.069; Variance 84.044; scale 0.274

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description	Pred. No.
1	256	100.0	123	33	W61383		Mouse HFGAN72 recepto	1.85e-18
2	256	100.0	130	30	W50158		Mouse hypocretin 35	1.85e-18
3	256	100.0	130	33	W61382		Rat HFGAN72 receptor	1.85e-18
4	256	100.0	130	30	W50157		Rat hypocretin 35	1.85e-18
5	256	100.0	131	33	W61381		Human HFGAN72 recepto	1.85e-18
6	77	30.1	363	23	W12414		Porcine complement in	8.08e+00
7	75	29.3	373	31	W56031		Mouse ICH-3	1.22e+01
8	75	29.3	373	13	R65767		Murine interleukin-1	1.22e+01
9	75	29.3	373	18	R98461		Murine ICE-ced-3 homo	1.22e+01
10	73	28.5	1876	29	W38757		Phosphatidyl inositol	1.85e+01
11	69	27.0	329	26	W29877		Lysophosphatidic acid	4.19e+01
12	66	25.8	102	1	P95679		Xenopus vgl protein f	7.64e+01
13	66	25.8	348	21	W01619		Human uridine diphosph	7.64e+01
14	65	25.4	14	30	W52512		Cyclic peptide of the	9.31e+01
15	65	25.4	78	5	P20020		Sequence of a foot an	9.31e+01
16	65	25.4	92	30	W53894		Fragment of chimeric	9.31e+01

17	65	25.4	396	33	W62598	Human 7-transmembrane	9.31e+01
18	64	25.0	45	23	W10105	Murine agouti signall	1.13e+02
19	64	25.0	45	23	W10106	Human agouti signall	1.13e+02
20	64	25.0	48	9	R45611	AG1 toxin.	1.13e+02
21	64	25.0	48	11	R60293	Calcium channel inhib	1.13e+02
22	64	25.0	48	8	R44209	A. aperta venom fract	1.13e+02
23	64	25.0	130	23	W10102	Human agouti signall	1.13e+02
24	64	25.0	131	38	W67568	Mouse agouti protein.	1.13e+02
25	64	25.0	131	23	W10101	Murine agouti signall	1.13e+02
26	64	25.0	131	33	W68617	Protein encoded by ag	1.13e+02
27	64	25.0	1167	27	W31504	Nematode toxin 167P p	1.13e+02
28	64	25.0	1167	20	W10653	Bacillus thuringiens	1.13e+02
29	64	25.0	1168	23	W16326	Nematocidal toxin 157	1.13e+02
30	63	24.6	583	30	W52118	Trypanosoma lymphocy	1.38e+02
31	62	24.2	34	10	R55088	Tarantula spider veno	1.68e+02
32	62	24.2	142	35	W76233	Bacterial periplasmic	1.68e+02
33	62	24.2	612	29	W46772	Protein of the specif	1.68e+02
34	62	24.2	1464	12	R66039	Human N-methyl-D-aspa	1.68e+02
35	62	24.2	1464	38	W85576	Human N-methyl-D-aspa	1.68e+02
36	62	24.2	1464	8	R42054	Glutamic acid recepto	1.68e+02
37	62	24.2	1464	18	R80970	Human excitatory amin	1.68e+02
38	62	24.2	1464	10	R55529	Human NMDA R2A recept	1.68e+02
39	62	24.2	1464	13	R44192	Rat NMDA receptor sub	1.68e+02
40	61	23.8	42	20	W12745	A-lineage conotoxin p	2.04e+02
41	61	23.8	42	23	W24890	Predatory cone snail	2.04e+02
42	61	23.8	253	25	W27333	Human membrane antige	2.04e+02
43	61	23.8	374	12	R62758	TctA sequence.	2.04e+02
44	61	23.8	439	38	W82645	Ehrlichia sp. extende	2.04e+02
45	61	23.8	3801	27	W31949	Human bg protein asso	2.04e+02

ALIGNMENTS

RESULT 1  
ID W61383 standard; Protein; 123 AA.  
AC W61383;

DT 02-OCT-1998 (first entry)  
DE Mouse HFGAN72 receptor protein.  
KW HFGAN72 receptor; eating disorders; renal disease; heart failure;  
KW sexual dysfunction; HIV; cancer; Parkinson's disease; urinary retention;  
KW osteoporosis; angina pectoris; myocardial infarction; psychotic disorder;  
KW neurological disorder.  
OS Mus sp.  
FH Key Location/Qualifiers  
FT Region 33..65  
FT /note= "Ligand 72A"  
FT /note= "Ligand 72B"

FT EP-849361-A2.  
FT 24-JUN-1998.  
FT 17-DEC-1997; 310216.  
FT 26-SEP-1997; US-939093.  
FT 17-DEC-1996; US-033604.  
FT 19-MAR-1997; US-820519.  
FT 02-JUL-1997; US-887382.  
FT (SMIK ) SMITHKLINE BEECHAM CORP.  
FT (SMIK ) SMITHKLINE BEECHAM PLC.  
FT Bergsma DJ, Brooks DP, Gellai M, Wilson S, Yanagisawa M;  
FT WPI; 98-324672/29.  
FT HFGAN72 receptor ligands - and corresponding DNA, agonists,  
FT antibodies, antagonists, etc.  
FT Claim 5; Fig 5; 35pp; English.  
FT The HFGAN72 receptor protein contains two ligands whose antagonists can be  
FT used for treating obesity, diabetes, anorexia nervosa, bulimia, cachexia,  
FT chronic renal failure, renal disease, congestive heart failure, impaired  
FT glucose tolerance and sexual dysfunction. The agonist is useful for  
FT treating anorexia nervosa, bulimia and cachexia. The HFGAN72 receptor  
FT ligand is useful for treating e.g. bacterial, fungal, protozoan and viral  
FT infections, particularly infections caused by HIV-1 or HIV-2, pain,  
FT cancers, anorexia nervosa, bulimia, cachexia, obesity, diabetes, asthma,  
FT Parkinson's disease, both acute and congestive heart failure,  
FT hypotension, hypertension, urinary retention, osteoporosis, angina  
FT pectoris, myocardial infarction, ulcers, asthma, allergies, benign

CC antagonists, in therapy, to detect Ab and to isolate cognate  
 CC receptors. Oligonucleotides based on H35 cDNA can be used to detect  
 CC the hyporetin gene or its RNA transcript, and as antisense agents  
 CC for inhibiting gene expression. H35 cDNA can also be used for  
 CC recombinant protein production. The Ab can be used to detect or  
 CC quantify hyporetin proteins and as a therapeutic inhibitor.  
 SQ Sequence 130 AA;

Query Match 100.0%; Score 256; DB 30; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 1.85e-18;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 33 qplpdcrrkqtcscrlyellhgaghaagiltl 65  
 Qy 1 QPLPDCRRKQTCSCRLYELLHGAGHAGHAGILTL 33

RESULT 5  
 ID W61381 standard; Protein; 131 AA.  
 AC W61381;  
 DT 02-OCT-1998 (first entry)  
 DE Human HFGAN72 receptor protein.  
 KW HFGAN72 receptor; eating disorders; renal disease; heart failure;  
 KW sexual dysfunction; HIV; cancer; Parkinson's disease; urinary retention;  
 KW osteoporosis; angina pectoris; myocardial infarction; psychotic disorder;  
 KW neurological disorder.  
 OS Homo sapiens.

FH Key Location/Qualifiers  
 FT Region 34..66  
 FT /note= "Ligand 72A"  
 FT Region 70..97  
 FT /note= "Ligand 72B"

EP-849361-A2.  
 PN 24-JUN-1998.  
 PD 17-DEC-1997; 310216  
 PF 26-SEP-1997; US-939093.  
 PR 17-DEC-1996; US-033604.  
 PR 19-MAR-1997; US-820519.  
 PR 02-JUL-1997; US-887382.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Bergsma DJ, Brooks DP, Geilal M, Wilson S, Yanagisawa M;  
 DR WPI; 98-324672/29.  
 DR N-PSDB; V28138.  
 DT HFGAN72 receptor ligands - and corresponding DNA, agonists,  
 PT antibodies, antagonists, etc.  
 PS Claim 2; Fig 2; 35pp; English.

CC The HFGAN72 receptor protein contains two ligands whose antagonists can  
 CC be used for treating obesity, diabetes, anorexia nervosa, bulimia,  
 CC cachexia, chronic renal failure, renal disease, congestive heart failure,  
 CC impaired glucose tolerance and sexual dysfunction. The agonist is  
 CC useful for treating anorexia nervosa, bulimia and cachexia. The HFGAN72  
 CC receptor ligand is useful for treating e.g. bacterial, fungal, protozoan  
 CC and viral infections, particularly infections caused by HIV-1 or HIV-2,  
 CC pain, cancers, anorexia nervosa, bulimia, cachexia, obesity, diabetes,  
 CC asthma, Parkinson's disease, both acute and congestive heart failure,  
 CC hypotension, hypertension, urinary retention, osteoporosis, angina  
 CC pectoris, myocardial infarction, ulcers, asthma, allergies, benign  
 CC prostatic hypertrophy, chronic renal failure, renal disease, impaired  
 CC glucose tolerance, sexual dysfunction and psychotic and neurological  
 CC disorders including anxiety, schizophrenia, manic depression, delirium,  
 CC dementia, severe mental retardation and dyskinesias such as Huntington's  
 CC disease or Gilles de la Tourette's syndrome.

SQ Sequence 131 AA;

Query Match 100.0%; Score 256; DB 33; Length 131;  
 Best Local Similarity 100.0%; Pred. No. 1.85e-18;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 34 qplpdcrrkqtcscrlyellhgaghaagiltl 66  
 Qy 1 QPLPDCRRKQTCSCRLYELLHGAGHAGHAGILTL 33

# RESULT 6

ID W12414 standard; Protein; 363 AA.  
 AC W12414;  
 DT 24-SEP-1997 (first entry)  
 DE Porcine complement inhibitor.  
 DE Porcine complement inhibitor.  
 KW porcine; pig; complement; inhibitor; organ transplantation;  
 KW analysis; promoter.  
 OS Sus scrofa.

PN WO9700951-A1.

PD 09-JAN-1997.

PE 19-JUN-1996; J01704.

PR 20-JUN-1995; JP-178254.

PA (NIME-) NIPPON MEAT PACKERS INC.

PA (NIHA-) NIPPON HAM KK.

PI Murakami H, Shigehisa T, Toyomura K;

DR WPI; 97-087378/08.

DR N-PSDB; T61098.

DT DNA encoding porcine complement inhibitor - useful in porcine organ

PT transplant to humans

PS Claim 3; Page 12-14; 20pp; Japanese.

CC This protein is a porcine complement inhibitor encoded by pmCpCDNA

CC (T61098). The DNA is useful for large scale production of

CC recombinant porcine complement inhibitor, which is useful for

CC porcine organ transplantation into humans. The DNA clone pmCpCDNA is

CC also useful in the analysis of the promoter region of porcine complement

CC inhibitor.

SQ Sequence 363 AA;

Query Match 30.1%; Score 77; DB 23; Length 363;

Best Local Similarity 39.3%; Pred. No. 8.08e-00;

Matches 11; Conservative 7; Mismatches 9; Indels 1; Gaps 1;

Db 98 plqaearrkacs-nlpdplngqvsypng 124

Qy 2 PLPDCRRKQTCSCRLYELLHGAGHAG 29

# RESULT 7

ID W56031 standard; Protein; 373 AA.

AC W56031;

DT 28-JUL-1998 (first entry)

DE Mouse ICH-3.

KW ICH-3; mouse; programmed cell death; septic shock; folliculogenesis;

KW interleukin converting enzyme; ICE; cysteine protease; cytokine;

KW maturation; apoptosis; sepsis; burn; trauma; immunogen.

OS Mus sp.

PN WO9806263-A1.

PD 19-FEB-1998.

PF 08-AUG-1997; U13898

PR 09-AUG-1996; US-023937.

PA (FISH/) FISHMAN J A.

PA (MIUR/) MIURA M.

PA (WANG/) WANG S.

PA (YUAN/) YUAN J.

PI Fishman JA, Miura M, Wang S, Yuan J;

DR WPI; 98-159183/14.

DR N-PSDB; V28526.

DT Transgenic knock-out mice containing new ICH-3 disrupted gene -

PT useful in, e.g. screening for compounds to treat septic shock and

PT defects in folliculogenesis

PS Example 1; Page 67.4-67.5; 99pp; English.

CC The present sequence represents mouse ICH-3 from the present invention.

CC The present invention describes: (1) a method for modulating programmed

CC cell death in a cell, comprising contacting the cell with modulating

CC amounts of ICH-3; (2) a method for promoting pro-interleukin (pro-IL)

CC 1-beta processing by a cell in the presence of IL-converting enzyme

CC (ICE), comprising contacting the cell with ICH-3 expressed under the

CC control of a CMV promoter; (3) a method for stimulating synthesis of

CC ICH-3 gene products in a cell, comprising contacting the cell with

CC stimulatory amounts of lipopolysaccharide (LPS); (4) a monoclonal or

CC polyclonal antibody that specifically binds to ICH-3; (5) a transgenic

CC non-human animal, comprising a disrupted ICH-3 gene, and its progeny,





```

RL INT. IMMUNOL. 9:869-876(1997).
DR EMBL; D70897; D1021310; -.
DR PFAM; PF00084; sushi; 4.
KW MEMBRANE.
SQ SEQUENCE 363 AA; 39692 MW; AD14F57A CRC32;

Query Match      30.1%; Score 77; DB 6; Length 363;
Best Local Similarity 39.3%; Pred. No. 7.54e-02;
Matches 11; Conservative 7; Mismatches 9; Indels 1; Gaps 1;

Db 98 PLQACRRKACS-NLPDPLNGVSPNG 124
Qy 2 PLPDCCRKQKTCSCRLYLHGAGNHAAG 29

RESULT 6
ID Q93473 PRELIMINARY; PRT; 752 AA.
AC Q93473; Q93693;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE F14B4.1 PROTEIN.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEAE; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIIDAE; PELODERINAE; CAENORHABDITIS.
[1]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
ST STEWARD C.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RX SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSOUCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR EMBL; 268219; E1349116; -.
SQ SEQUENCE 683 AA; 77437 MW; B9B00EA2 CRC32;

Query Match      28.5%; Score 73; DB 5; Length 683;
Best Local Similarity 50.0%; Pred. No. 3.80e-01;
Matches 7; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

Db 382 ECCQKKECECR-FE 394
Qy 5 DCCROKTCSCRLYE 18

RESULT 8
ID Q24209 PRELIMINARY; PRT; 1876 AA.
AC Q24209;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOSPHOINOSITIDE 3-KINASE.
GN CPK.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-OREGON R;
RC MEDLINE; 96278830.
RX MOLZ L.M., CHEN Y.W., HIRANO M., WILLIAMS L.T.;
RT "Cpk is a novel class of Drosophila PtdIns 3-kinase containing a C2
RT domain.";
RL J. BIOL. CHEM. 271:13892-13899(1996).
DR EMBL; U52192; G1272420; -.
DR FLYBASE; FBgn0015278; P13K68D.
DR PFAM; PF00168; C2; 1.
DR PFAM; PF00454; P13_P14_kinase; 1.
DR PFAM; PF00613; P13Ka; 1.
DR PFAM; PF00787; PX; 1.
DR PFAM; PF00792; P13K_C2; 1.
DR PFAM; PF00794; P13K_rbd; 1.
SQ SEQUENCE 1876 AA; 210504 MW; DF107ECF CRC32;

Query Match      28.5%; Score 73; DB 5; Length 1876;
Best Local Similarity 41.7%; Pred. No. 3.80e-01;
Matches 10; Conservative 5; Mismatches 8; Indels 1; Gaps 1;

```

```

RL INT. IMMUNOL. 9:869-876(1997).
DR EMBL; D70897; D1021310; -.
DR PFAM; PF00084; sushi; 4.
KW MEMBRANE.
SQ SEQUENCE 363 AA; 39692 MW; AD14F57A CRC32;

Query Match      30.1%; Score 77; DB 6; Length 363;
Best Local Similarity 39.3%; Pred. No. 7.54e-02;
Matches 11; Conservative 7; Mismatches 9; Indels 1; Gaps 1;

Db 98 PLQACRRKACS-NLPDPLNGVSPNG 124
Qy 2 PLPDCCRKQKTCSCRLYLHGAGNHAAG 29

RESULT 6
ID Q93473 PRELIMINARY; PRT; 752 AA.
AC Q93473; Q93693;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE F14B4.1 PROTEIN.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEAE; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIIDAE; PELODERINAE; CAENORHABDITIS.
[1]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
ST STEWARD C.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RX SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSOUCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
[3]
RN SEQUENCE FROM N.A.
RA SIMS M.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; 281078; E1346430; -.
DR EMBL; 275535; E1346430; JOINED.
DR EMBL; 275535; E1345510; -.
DR EMBL; 281078; E1345510; JOINED.
DR PROSITE; PS00010; ASX-HYDROXYL; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01209; LDLRA_1; 2.
KW GLYCOPROTEIN; EGF-LIKE DOMAIN.
SQ SEQUENCE 752 AA; 84505 MW; D060D4DA CRC32;

Query Match      28.9%; Score 74; DB 5; Length 752;
Best Local Similarity 56.3%; Pred. No. 2.55e-01;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 245 QPLBECDTQGNCKRL 260
Qy 1 QPLPDCCRKQKTCSCRL 16

RESULT 7
ID Q22187 PRELIMINARY; PRT; 683 AA.
AC Q22187;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

```

```

RL  elegans."
RL  NATURE 368:32-38(1994).
RN  [2]
RC  SEQUENCE FROM N.A.
RC  STRAIN-BRISTOL N2.
RA  GATTUNG S., GOELA D., WILSON R.;
RA  SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN-BRISTOL N2.
RA  WATERSTON R.;
RL  SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR  EMBL: AF003739; G2105488;
DR  PROSITE: PS00627; GHMP_KINASES_ATP; 1.
DR  PFAM: PF00288; GHMP_kinases; 1.
SQ  SEQUENCE 426 AA; 47290 MW; 7CBEF743 CRC32;

Query Match      27.3%; Score 70; DB 5; Length 426;
Best Local Similarity 48.1%; Pred. No. 1.23e+00;
Matches 13; Conservative 4; Mismatches 7; Indels 3; Gaps 3;

Db  361 ECSCRELDEICRLY-LDHGALGARLTG 386
      :| | | | | | | | | | | | | | | |
QY  5 DC-CROKTCSCRLYELLHGA-GNHAAG 29

RESULT 13
ID  Q91493      PRELIMINARY;      PRT;      883 AA.
AC  Q91493;
DT  01-NOV-1996 (TREMBLREL. 01, CREATED)
DT  01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT  01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE  DYSTROPHIN (FRAGMENT).
GN  DYSTROPHIN.
OS  TORPEDO CALIFORNICA (PACIFIC ELECTRIC RAY).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; CHONDRICHTHYES;
OC  ELASMOBRANCHII; RAJIFORMES; TORPEDINOIDEI; TORPEDINIDAE; TORPEDO.
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE: 92064638.
RA  YEADON J.E., LIN H., DYER S.M., BURDEN S.J.;
RT  "Dystrophin is a component of the subsynaptic membrane."
RL  J. CELL BIOL. 115:1069-1076(1991).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE: 92291902.
RA  RAVIN A.J., DYER S.M., YEADON J.E., BURDEN S.J.;
RT  "Multiple dystrophin isoforms are associated with the postsynaptic
RT  membrane of Torpedo electric organ."
RL  J. PHYSIOL. (PARIS) 0:131-133(1991).
DR  EMBL: M37645; G397971;
DR  PROSITE: PS01159; WW_DOMAIN_1; 1.
DR  PFAM: PF00397; WW_rsp5_WWP; 1.
DR  PFAM: PF00435; spectrin; 2.
DR  PFAM: PF00569; Z2; 1.
FT  NON_TER 1
SQ  SEQUENCE 883 AA; 101397 MW; 70879876 CRC32;

Query Match      27.3%; Score 70; DB 13; Length 883;
Best Local Similarity 61.5%; Pred. No. 1.23e+00;
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db  826 LLHGVGSQTSGIL 838
      | | | | | | | | | | | | | | | |
QY  19 LLHGAGNHAAGIL 31

RESULT 14
ID  O68770      PRELIMINARY;      PRT;      1135 AA.
AC  O68770;
DT  01-AUG-1998 (TREMBLREL. 07, CREATED)
DT  01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT  01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

```

```

DE  DNA POLYMERASE III ALPHA SUBUNIT HOMOLOG.
OS  YERSINIA PESTIS.
OG  PLASMID PMT1.
OC  BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC  YERSINIA.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-KIM;
RA  HU P., ELLIOTT J., MCCREADY P., SKOWRONSKI E., GARNES J.,
RA  KOBAYASHI A., CARRANO A.V., BRUBAKER R., GARCIA E.;
RL  SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR  EMBL: AF053947; G2996335;
KW  PLASMID.
SQ  SEQUENCE 1135 AA; 126286 MW; 4EED77D7 CRC32;

Query Match      27.3%; Score 70; DB 2; Length 1135;
Best Local Similarity 47.4%; Pred. No. 1.23e+00;
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db  474 ACKLQSLMRGFRHAAGMI 492
      :| | | | | | | | | | | | | | | |
QY  13 SCRLYELLHGAGNHAAGIL 31

RESULT 15
ID  O52422      PRELIMINARY;      PRT;      163 AA.
AC  O52422;
DT  01-JUN-1998 (TREMBLREL. 06, CREATED)
DT  01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT  01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE  SHIKIMATE DEHYDROGENASE (FRAGMENT).
GN  AROE.
OS  NEISSERIA MENINGITIDIS.
OC  BACTERIA; PROTEOBACTERIA; BETA SUBDIVISION; NEISSERIACEAE; NEISSERIA.
RN  [1]
RP  SEQUENCE FROM N.A.
RA  MAIDEN M.C.J., BYGRAVES J.A., FEIL E., MORELLI G., RUSSELL J.E.,
RA  URWIN R., ZHANG Q., ZHOU J., ZURTH K., CAUGANT D.A., FEAVERS I.M.,
RA  ACHTMAN M., SPRATT B.G.;
RL  SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR  EMBL: AF037797; G2745813;
FT  NON_TER 1
FT  NON_TER 163
SQ  SEQUENCE 163 AA; 17302 MW; ED405667 CRC32;

Query Match      27.0%; Score 69; DB 2; Length 163;
Best Local Similarity 34.6%; Pred. No. 1.82e+00;
Matches 9; Conservative 7; Mismatches 9; Indels 1; Gaps 1;

Db  90 LPAYSVREIFRNCRLAYDMVYDAAQA 115
      | | | | | | | | | | | | | | | |
QY  3 LPDCRQKTCSCRL-YELLHGAGNHA 27

Search completed: Fri Aug 20 21:08:05 1999
Job time : 41 secs.

```



\*\*\*\*\*

W P E R F H (TW)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:09:47 1999; MasPar time 4.51 Seconds  
132.123 Million cell updates/sec

Tabular output not generated.

Title: >US-08-938-548B-9  
Description: (1-28) from US08938548B.pep  
Perfect Score: 201  
Sequence: 1 RFGPPGLOGRLQRLQANGNHAAGILTM 28  
Scoring table: PAM 150  
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 22.192; Variance 88.282; scale 0.251

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	201	100.0	123	33	Mouse HFGAN72 recepto	2.26e-11
2	201	100.0	130	33	Rat HFGAN72 receptor	2.26e-11
3	201	100.0	130	30	Mouse hypocretin 35.	2.26e-11
4	201	100.0	130	30	Rat hypocretin 35.	2.26e-11
5	190	94.5	131	33	Human HFGAN72 recepto	2.79e-10
6	68	33.8	2192	23	LexA/NUMA fusion prot	6.37e+01
7	68	33.8	2272	23	GALA/NUMA fusion p	6.37e+01
8	66	32.8	10	4	"p33" N-terminal (2).	9.27e+01
9	65	32.3	716	18	Retinoid X receptor i	1.12e+02
10	65	32.3	1464	35	Human recombinant col	1.12e+02
11	65	32.3	2509	24	Protein (OA-519) cros	1.12e+02
12	64	31.8	256	9	Deduced sequence of h	1.34e+02
13	64	31.8	256	15	MY17 preproPR-3.	1.34e+02
14	64	31.8	549	16	KM31-7 precursor.	1.34e+02
15	64	31.8	551	38	Human KM-102-derived	1.34e+02
16	64	31.8	1477	13	S. cerevisiae scaur2R	1.34e+02

17	64	31.8	1477	22	W10424	Saccharomyces cerevis	1.34e+02
18	63	31.3	550	29	W45513	Ecdysone receptor lig	1.62e+02
19	63	31.3	550	29	W39139	Drosophila ecdysone r	1.62e+02
20	63	31.3	746	28	W33655	Modified ecdysone rec	1.62e+02
21	63	31.3	746	28	W33654	Modified ecdysone rec	1.62e+02
22	63	31.3	878	3	R13793	Ecdysone receptor.	1.62e+02
23	63	31.3	878	6	R32889	DHR23alpha protein.	1.62e+02
24	63	31.3	1041	28	W33656	Modified ecdysone rec	1.62e+02
25	63	31.3	1841	26	W22605	Tyrosine synthase OR	1.62e+02
26	63	31.3	4630	23	W19629	Streptomyces venezuel	1.62e+02
27	62	30.8	684	24	W26327	Human alpha-1 collage	1.94e+02
28	62	30.8	1694	29	W40109	Human alpha-6(IV) col	1.94e+02
29	61	30.3	32	1	R00579	New polypeptide based	2.33e+02
30	61	30.3	248	1	R06331	Human alveolar surfac	2.33e+02
31	61	30.3	248	3	P60441	Plasmod PAspc-SV(10)	2.33e+02
32	61	30.3	248	2	P70662	35kd pulmonary surfac	2.33e+02
33	61	30.3	248	1	P80694	Sequence deduced from	2.33e+02
34	61	30.3	248	1	P82980	Sequence deduced from	2.33e+02
35	61	30.3	248	1	R04215	Human 32K ASP encoded	2.33e+02
36	61	30.3	248	3	P60666	Genomic sequence of h	2.33e+02
37	61	30.3	248	3	P60665	Sequence of human alv	2.33e+02
38	61	30.3	248	2	R05091	Vector PSP 35K-1A-10	2.33e+02
39	61	30.3	271	3	P60661	Genomic sequence of h	2.33e+02
40	61	30.3	271	1	R04212	Human 32K alveolar su	2.33e+02
41	61	30.3	271	1	R04217	Human 32K ASP encoded	2.33e+02
42	61	30.3	381	38	W81135	Human 3-hydroxyisobut	2.33e+02
43	61	30.3	495	28	W39747	Human marcosR protein	2.33e+02
44	61	30.3	520	28	W39748	Human marcosR protein	2.33e+02
45	61	30.3	1503	33	W48845	Human receptor tyrosi	2.33e+02

#### ALIGNMENTS

RESULT 1  
ID W61383 standard; Protein; 123 AA.  
AC W61383;  
DT 02-OCT-1998 (first entry)  
DE Mouse HFGAN72 receptor protein.  
KW HFGAN72 receptor; eating disorders; renal disease; heart failure;  
KW sexual dysfunction; HIV; cancer; Parkinson's disease; urinary retention;  
KW osteoporosis; angina pectoris; myocardial infarction; psychotic disorder;  
KW neurological disorder.  
OS Mus sp.  
FH Key Location/Qualifiers  
FT Region 33..65  
FT /note= "Ligand 72A"  
FT 69..96  
FT /note= "Ligand 72B"  
FN EP-849361-A2.  
PD 24-JUN-1998.  
PF 17-DEC-1997; 310216.  
PR 26-SEP-1997; US-939093.  
PR 17-DEC-1996; US-033604.  
PR 19-MAR-1997; US-820519.  
PR 02-JUL-1997; US-887382.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PI Bergsma DJ, Brooks DP, Gellai M, Wilson S, Yanagisawa M;  
DR WPI: 98-324672/29.  
PT HFGAN72 receptor ligands - and corresponding DNA, agonists,  
PT antibodies, antagonists, etc.  
PS Claim 5; Fig 5; 35pp; English.  
CC The HFGAN72 receptor protein contains two ligands whose antagonists can be  
CC used for treating obesity, diabetes, anorexia nervosa, bulimia, cachexia,  
CC chronic renal failure, renal disease, congestive heart failure, impaired  
CC glucose tolerance and sexual dysfunction. The agonist is useful for  
CC treating anorexia nervosa, bulimia and cachexia. The HFGAN72 receptor  
CC ligand is useful for treating e.g. Bacterial, fungal, protozoan and viral  
CC infections, particularly infections caused by HIV-1 or HIV-2, pain,  
CC cancers, anorexia nervosa, bulimia, cachexia, obesity, diabetes, asthma,  
CC Parkinson's disease, both acute and congestive heart failure,  
CC hypotension, hypertension, urinary retention, osteoporosis, angina  
CC pectoris, myocardial infarction, ulcers, asthma, allergies, benign

CC antagonists, in therapy, to detect Ab and to isolate cognate  
 CC receptors. Oligonucleotides based on H35 cDNA can be used to detect  
 CC the hypocretin gene or its RNA transcript, and as antisense agents  
 CC for inhibiting gene expression. H35 cDNA can also be used for  
 CC recombinant protein production. The Ab can be used to detect or  
 CC quantify hypocretin proteins and as a therapeutic inhibitor.  
 SQ Sequence 130 AA;

Query Match 100.0%; Score 201; DB 30; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 2.26e-11;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 69 rpgpglqrlqlqanqhaagiltm 96  
 QY 1 RFGPPGLQRLQLQANQHAAGILTM 28

RESULT 5  
 ID W61381 standard; Protein; 131 AA.  
 AC W61381;  
 DT 02-OCT-1998 (first entry)  
 DE Human HFGAN72 receptor protein.  
 KW HFGAN72 receptor; eating disorders; renal disease; heart failure;  
 KW sexual dysfunction; HIV; cancer; Parkinson's disease; urinary retention;  
 KW osteoporosis; angina pectoris; myocardial infarction; psychotic disorder;  
 KW neurological disorder.  
 OS Homo sapiens.

FH Key Location/Qualifiers  
 FT Region 34..66  
 FT /note= "Ligand 72A"  
 FT Region 70..97  
 FT /note= "Ligand 72B"

PN EP-849361-A2.  
 PD 24-JUN-1998.  
 PR 17-DEC-1997; 310216.  
 PR 26-SEP-1997; US-939093.  
 PR 17-DEC-1996; US-033604.  
 PR 19-MAR-1997; US-820519.  
 PR 02-JUL-1997; US-887382.  
 PA (SMK ) SMITHKLINE BEECHAM CORP.  
 PA (SMK ) SMITHKLINE BEECHAM PLC.  
 PI Bergsma DJ, Brooks DP, Gellai M, Wilson S, Yanagisawa M;  
 DR WPI; 98-324672/29.  
 DR N-PSDB; V28138.  
 PT HFGAN72 receptor ligands - and corresponding DNA, agonists,  
 PT antibodies, antagonists, etc.  
 PS Claim 2; Fig 2; 35pp; English.

CC The HFGAN72 receptor protein contains two ligands whose antagonists can  
 CC be used for treating obesity, diabetes, anorexia nervosa, bulimia,  
 CC cachexia, chronic renal failure, renal disease, congestive heart failure,  
 CC impaired glucose tolerance and sexual dysfunction. The agonist is  
 CC useful for treating anorexia nervosa, bulimia and cachexia. The HFGAN72  
 CC receptor ligand is useful for treating e.g. bacterial, fungal, protozoan  
 CC and viral infections, particularly infections caused by HIV-1 or HIV-2,  
 CC pain, cancers, anorexia nervosa, bulimia, cachexia, obesity, diabetes,  
 CC asthma, Parkinson's disease, both acute and congestive heart failure,  
 CC hypotension, hypertension, urinary retention, osteoporosis, angina  
 CC pectoris, myocardial infarction, ulcers, asthma, allergies, benign  
 CC prostatic hypertrophy, chronic renal failure, renal disease, impaired  
 CC glucose tolerance, sexual dysfunction and psychotic and neurological  
 CC disorders including anxiety, schizophrenia, manic depression, delirium,  
 CC dementia, severe mental retardation and dyskinesias such as Huntington's  
 CC disease or Gilles de la Tourette's syndrome.  
 SQ Sequence 131 AA;

Query Match 94.5%; Score 190; DB 33; Length 131;  
 Best Local Similarity 92.9%; Pred. No. 2.79e-10;  
 Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 70 rsgpglqrlqlqasqnaagiltm 97  
 QY 1 RFGPPGLQRLQLQANQHAAGILTM 28

RESULT 6  
 ID W21732 standard; Protein; 2192 AA.  
 AC W21732;  
 DT 01-OCT-1997 (first entry)  
 DE LexA/NumA fusion protein.  
 KW NIP-1; NIP-2; NuMA; nuclear mitotic apparatus; NuMA interacting protein;  
 KW cell division; proliferation; antibody; Ab; detection;  
 KW malignant cell growth.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT domain 1..87  
 FT /label= LexA\_DNA\_binding\_domain  
 FT peptide 88..94  
 FT /label= Polylinker  
 FT protein 95..2192  
 FT /label= Residues\_18-2116\_of\_NuMA  
 FT region 285..1784  
 FT /label= Coiled\_coil\_region

PN W09640917-Al.  
 PD 19-DEC-1996.  
 PF 07-JUN-1996; U09504.  
 PR 07-JUN-1995; US-478408.  
 PA (UYA ) UNIV YALE.  
 PI McPherson SMG, Snyder MP;  
 DR WPI; 97-077270/07.  
 DR N-PSDB; T77783.  
 PT New nucleic acid encoding nuclear mitotic appts. interacting  
 PT proteins - useful for modulating cell division and proliferation and  
 PT in diagnosis  
 PS Claim 15; Page 42-50; 78pp; English.  
 CC The sequences given in W21731-32 represent fusion proteins which contain  
 CC NuMA (nuclear mitotic apparatus). The fusion proteins were used in  
 CC the identification of NuMA interacting proteins (NIP's) (see also  
 CC W21729-30). Compounds which interfere with the interaction of NuMA  
 CC with a known NIP are used to modulate cell division and/or proliferation.  
 CC Ab, raised conventionally using NIP-1 or -2 as immunogen, are used to  
 CC detect NIP (or their complexes) and to block their activity for  
 CC diagnostic or therapeutic use, e.g. to detect defective NuMA or NIP  
 CC which may be markers for aberrant (including malignant) cell growth  
 CC (which can also be detected by nucleic acid sequencing). Also where  
 CC malignancy is related to defects in NuMA or NIP, it can be treated by  
 CC administration of the appropriate functional protein.  
 SQ Sequence 2192 AA;

Query Match 33.8%; Score 68; DB 23; Length 2192;  
 Best Local Similarity 50.0%; Pred. No. 6.37e+01;  
 Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 838 ragrkqlearlqqlgeah 855  
 QY 1 RFGPPGLQRLQLQAN 18

RESULT 7  
 ID W21731 standard; Protein; 2272 AA.  
 AC W21731;  
 DT 01-OCT-1997 (first entry)  
 DE GAL4/HA/NuMA fusion protein.  
 KW NIP-1; NIP-2; NuMA; nuclear mitotic apparatus; NuMA interacting protein;  
 KW cell division; proliferation; antibody; Ab; detection;  
 KW malignant cell growth.  
 OS Homo sapiens.

FH Key Location/Qualifiers  
 FT domain 1..147  
 FT /label= GAL4\_DNA\_binding\_domain  
 FT peptide 148..174  
 FT /label= Hemagglutinin epitope  
 FT protein 175..2272  
 FT /label= Residues\_18-2116\_of\_NuMA  
 FT region 365..1864  
 FT /label= Coiled\_coil\_region  
 PN W09640917-Al.



\*\*\*\*\*

W P S R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:12:58 1999; MasPar time 1.89 Seconds

Tabular output not generated. 150.169 Million cell updates/sec

Title: >US-08-938-548B-9  
Description: (1-28) from US08938548B.pep  
Perfect Score: 201  
Sequence: 1 RGPPLGQRLQRLQALQANGNHAAGILTM 28

Scoring table: PAM 150  
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 20.582; Variance 81.576; scale 0.252

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	67	33.3	442	2	US-08-363- Sequence 4, Applicatio	2.72e+01
2	67	33.3	442	2	US-08-363- Sequence 11, Applicati	2.72e+01
3	67	33.3	445	2	US-08-363- Sequence 5, Applicati	2.72e+01
4	67	33.3	445	2	US-08-363- Sequence 12, Applicati	2.72e+01
5	66	32.8	10	3	PCT-US91-0 Sequence 3, Applicatio	3.30e+01
6	65	32.3	716	3	PCT-US91-1 Sequence 4, Applicatio	4.00e+01
7	65	32.3	2509	1	US-08-469- Sequence 10, Applicati	4.00e+01
8	63	31.3	1841	2	US-08-804- Sequence 6, Applicatio	5.87e+01
9	62	30.8	1694	1	US-08-494- Sequence 2, Applicatio	7.10e+01
10	61	30.3	381	2	US-08-858- Sequence 3, Applicatio	8.57e+01
11	61	30.3	381	2	US-08-858- Sequence 1, Applicatio	8.57e+01
12	60	29.9	199	3	PCT-US93-0 Sequence 2, Applicatio	1.03e+02
13	60	29.9	199	3	US-07-949- Sequence 4, Applicatio	1.03e+02
14	60	29.9	199	1	US-08-115- Sequence 2, Applicatio	1.03e+02
15	60	29.9	199	1	US-08-792- Sequence 8, Applicatio	1.03e+02
16	60	29.9	199	2	US-08-814- Sequence 4, Applicatio	1.03e+02
17	60	29.9	199	1	US-08-017- Sequence 2, Applicatio	1.03e+02
18	60	29.9	199	1	US-07-941- Sequence 4, Applicatio	1.03e+02
19	60	29.9	296	1	PCT-US93-0 Sequence 14, Applicati	1.03e+02
20	60	29.9	296	3	PCT-US93-0 Sequence 4, Applicatio	1.03e+02
21	60	29.9	296	3	PCT-US94-1 Sequence 14, Applicati	1.03e+02
22	60	29.9	296	1	US-08-115- Sequence 4, Applicatio	1.03e+02
23	60	29.9	296	1	US-07-941- Sequence 4, Applicatio	1.03e+02

24	60	29.9	296	1	US-08-165- Sequence 14, Applicati	1.03e+02
25	60	29.9	296	1	US-07-745- Sequence 14, Applicati	1.03e+02
26	60	29.9	300	2	US-08-794- Sequence 2, Applicatio	1.03e+02
27	60	29.9	405	1	US-07-688- Sequence 14, Applicati	1.03e+02
28	60	29.9	405	3	PCT-US91-0 Sequence 14, Applicati	1.03e+02
29	60	29.9	528	2	US-08-363- Sequence 14, Applicati	1.03e+02
30	60	29.9	530	2	US-08-363- Sequence 8, Applicatio	1.03e+02
31	60	29.9	530	2	US-08-363- Sequence 10, Applicati	1.03e+02
32	60	29.9	530	2	US-08-363- Sequence 3, Applicatio	1.03e+02
33	60	29.9	530	2	US-08-363- Sequence 2, Applicatio	1.03e+02
34	60	29.9	530	2	US-08-363- Sequence 9, Applicatio	1.03e+02
35	59	29.4	116	1	US-08-687- Sequence 3, Applicatio	1.25e+02
36	59	29.4	116	2	US-08-816- Sequence 3, Applicatio	1.25e+02
37	59	29.4	176	3	PCT-US96-1 Sequence 6, Applicatio	1.25e+02
38	59	29.4	176	2	US-08-659- Sequence 6, Applicatio	1.25e+02
39	59	29.4	430	2	US-08-318- Sequence 16, Applicati	1.25e+02
40	59	29.4	588	2	US-08-620- Sequence 2, Applicatio	1.25e+02
41	59	29.4	3724	2	US-08-804- Sequence 10, Applicati	1.25e+02
42	58	28.9	199	4	5215895-2 Patent No. 5215895.	1.50e+02
43	58	28.9	226	4	5498600-2 Patent No. 5498600.	1.50e+02
44	58	28.9	241	4	5175255-8 Patent No. 5175255.	1.50e+02
45	58	28.9	1367	1	US-07-946- Sequence 4, Applicatio	1.50e+02

#### ALIGNMENTS

RESULT 1	ID	US-08-363-255-4	STANDARD;	PRT;	442 AA.
XX	XX	xxxxxx			
XX	XX				
XX	XX				
XX	XX				
XX	XX				
CC	CC	Sequence 4, Application US/08363255			
CC	CC	Sequence 4, Application US/08363255			
CC	CC	Patent No. 5783386			
CC	CC	GENERAL INFORMATION:			
CC	CC	APPLICANT: JACOBS, JR., WILLIAM R.			
CC	CC	APPLICANT: BLOOM, BAREY R.			
CC	CC	APPLICANT: COLLINS, DESMOND M.			
CC	CC	APPLICANT: de LISLE, GEOFFREY W.			
CC	CC	APPLICANT: PASCOPELLA, LISA			
CC	CC	APPLICANT: KAWAKAMI, RIKU P.			
CC	CC	TITLE OF INVENTION: MYCOBACTERIA VIRULENCE FACTORS AND A			
CC	CC	TITLE OF INVENTION: NOVEL METHOD FOR THEIR IDENTIFICATION			
CC	CC	NUMBER OF SEQUENCES: 14			
CC	CC	CORRESPONDENCE ADDRESS:			
CC	CC	ADDRESSEE: MORRISON & FOERSTER			
CC	CC	STREET: 755 Page Mill Road			
CC	CC	CITY: Palo Alto			
CC	CC	STATE: California			
CC	CC	COUNTRY: USA			
CC	CC	ZIP: 94304-1018			
CC	CC	COMPUTER READABLE FORM:			
CC	CC	MEDIUM TYPE: Floppy disk			
CC	CC	COMPUTER: IBM PC compatible			
CC	CC	OPERATING SYSTEM: PC-DOS/MS-DOS			
CC	CC	SOFTWARE: Patentin Release #1.0, Version #1.30			
CC	CC	CURRENT APPLICATION DATA:			
CC	CC	APPLICATION NUMBER: US/08/363,255			
CC	CC	FILING DATE: 23-DEC-1994			
CC	CC	CLASSIFICATION: 435			
CC	CC	ATTORNEY/AGENT INFORMATION:			
CC	CC	NAME: MONROY, GLADYS H.			
CC	CC	REGISTRATION NUMBER: 32,430			
CC	CC	REFERENCE/DOCKET NUMBER: 25237-20002.22			
CC	CC	TELECOMMUNICATION INFORMATION:			
CC	CC	TELEPHONE: (415) 813-5600			
CC	CC	TELEFAX: (415) 494-0792			
CC	CC	INFORMATION FOR SEQ ID NO: 4:			
CC	CC	SEQUENCE CHARACTERISTICS:			

CC	CITY:	New York
CC	STATE:	New York
CC	COUNTRY:	USA
CC	ZIP:	10022-6250
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	Floppy disk
CC	COMPUTER:	IBM PC compatible
CC	OPERATING SYSTEM:	PC-DOS/MS-DOS
CC	SOFTWARE:	PatentIn Release #1.0, Version #1.25
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	PCT/US91/04588
CC	FILING DATE:	19910627
CC	CLASSIFICATION:	530
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	US 07/544,862
CC	FILING DATE:	27-JUN-1990
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME:	Haley Jr., James F.
CC	REGISTRATION NUMBER:	27,794
CC	REFERENCE/DOCKET NUMBER:	B129CIP
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE:	212-715-0600
CC	TELEFAX:	212-715-0674
CC	TELEX:	14-8367
CC	INFORMATION FOR SEQ ID NO:	3:
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH:	10 amino acids
CC	TYPE:	AMINO ACID
CC	TOPOLOGY:	unknown
CC	MOLECULE TYPE:	peptide
SQ	SEQUENCE	10 AA; 1169 MW; 376 CN;
Query Match	32.8%;	Score 66; DB 3; Length 10;
Best Local Similarity	88.9%;	Pred. No. 3.30e+01;
Matches	8; Conservative	1; Mismatches 0; Indels 0; Gaps 0
Db	1 GLEGRLQLR 9	
	:	
Qy	6 GLOGRLQLR 14	
RESULT	6	
ID	PCT-US95-16311-4	STANDARD; PRT; 716 AA.
XX	xxxxxx	
AC		
DT		
XX		
DE		
XX		
CC	Sequence 4, Application PC/TUS9516311	
CC	Sequence 4, Application PC/TUS9516311	
CC	GENERAL INFORMATION:	
CC	APPLICANT:	Moore, David
CC	APPLICANT:	Seol, Wongi
CC	APPLICANT:	Choi, Hwang-Sik
CC	TITLE OF INVENTION:	RETINOID X RECEPTOR-INTERACTING
CC	TITLE OF INVENTION:	POLYPEPTIDES AND RELATED MOLECULES AND METHODS
CC	NUMBER OF SEQUENCES:	17
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE:	Fish & Richardson P.C.
CC	STREET:	225 Franklin Street, Suite 3100
CC	CITY:	Boston
CC	STATE:	MA
CC	COUNTRY:	USA
CC	ZIP:	02110-2804
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	Floppy disk
CC	COMPUTER:	IBM PC compatible
CC	OPERATING SYSTEM:	PC-DOS/MS-DOS
CC	SOFTWARE:	PatentIn Release #1.0, Version #1.30
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	PCT/US95/16311
CC	FILING DATE:	

```

Query Match          31.3%; Score 63; DB 2; Length 1841;
Best Local Similarity 33.3%; Pred. No. 5.87e+01;
Matches            8; Conservative      8; Mismatches    8; Indels    0; Gaps    0;

Db   1513 RDTPAALAAHLAELLATARDHGP 1536
QY   1 RPGLPGLQLRLLQANGNHAAG 24
| :||::|| || :::||
| :||::|| || :::||

RESULT           9
ID US-08-494-168-2 STANDARD; PRT; 1694 AA.
XX AC xxxxxx
XX XX
XX DT
XX DE
XX DE
XX DE
CC CC Sequence 2, Application US/08494168
CC CC Sequence 2, Application US/08494168
CC CC Patent No. 5731192
CC CC GENERAL INFORMATION:
CC CC APPLICANT: Readers, Stephen T.
CC CC APPLICANT: Zhou, Jing
CC CC TITLE OF INVENTION: Collagen COLA6: Gene, Protein and Method
CC CC TITLE OF INVENTION: of Detecting Collagen Deficiency
CC CC NUMBER OF SEQUENCES: 10
CC CC CORRESPONDENCE ADDRESS:
CC CC ADDRESSEE: Foley & Lardner
CC CC STREET: 3000 K Street, N.W., Suite 500
CC CC CITY: Washington, D.C.
CC CC COUNTRY: USA
CC CC ZIP: 20007-5109
CC CC COMPUTER READABLE FORM:
CC CC MEDIUM TYPE: Floppy disk
CC CC COMPUTER: IBM PC compatible
CC CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CC CURRENT APPLICATION DATA:
CC CC APPLICATION NUMBER: US/08/494,168
CC CC FILING DATE:
CC CC CLASSIFICATION: 435
CC CC PRIOR APPLICATION DATA:
CC CC APPLICATION NUMBER: US 08/112,465
CC CC FILING DATE: 27-AUG-1993
CC CC ATTORNEY/AGENT INFORMATION:
CC CC NAME: SAXE, Bernhard D.
CC CC REGISTRATION NUMBER: 28,665
CC CC REFERENCE/DOCKET NUMBER: 40397/104/BABR
CC CC TELECOMMUNICATION INFORMATION:
CC CC TELEPHONE: (202)672-5300
CC CC TELEFAX: (202)672-5399
CC CC TELEX: 904136
CC CC INFORMATION FOR SEQ ID NO: 2:
CC CC SEQUENCE CHARACTERISTICS:
CC CC LENGTH: 1694 amino acids
CC CC TYPE: amino acid
CC CC TOPOLOGY: linear
CC CC MOLECULE TYPE: protein
SQ SEQUENCE 1694 AA; 165283 MW; 14956750 CN;

Query Match          30.8%; Score 62; DB 1; Length 1694;
Best Local Similarity 88.9%; Pred. No. 7.10e+01;
Matches            8; Conservative      0; Mismatches    1; Indels    0; Gaps    0;

Db   1028 RPLGPPGLPG 1036
QY   1 RPLGPPGLQG 9
||||||| |
||||||| |

RESULT           10
ID US-08-858-052-3 STANDARD; PRT; 381 AA.
XX AC xxxxxx

```

CC	CLASSIFICATION: 530
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: Meindert, M. C.
CC	REGISTRATION NUMBER: 31.544
CC	REFERENCE/DOCKET NUMBER: 5174BPCT
CC	INFORMATION FOR SEQ ID NO: 4:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 199 amino acids
CC	TYPE: amino acid
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: protein
CC	SEQUENCE 199 AA; 21429 MW; 188641 CN;
CC	Query Match 29.9%; Score 60; DB 1; Length 199;
CC	Best Local Similarity 70.08; Pred. No. 1.03e+02;
CC	Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
DB	129 LQARLDRLR 138
QY	7 LQGRLQRLQ 16
CC	RESULT 14
XX	US-08-115-680-2 STANDARD: PRT; 199 AA.
AC	xxxxxx
XX	
DT	
XX	
XX	Sequence 2, Application US/08115680
XX	
CC	Sequence 2, Application US/08115680
CC	Patent No. 5437863
CC	GENERAL INFORMATION:
CC	APPLICANT: Williams, David A.
CC	APPLICANT: Clark, Steven C.
CC	TITLE OF INVENTION: Method of Treating Cell Damage or
CC	TITLE OF INVENTION: Depletion
CC	NUMBER OF SEQUENCES: 4
CC	CORRESPONDENCE ADDRESSES:
CC	ADDRESSEE: Howson and Howson
CC	STREET: Spring House Corporate Cntr, P.O. Box 457
CC	CITY: Spring House
CC	STATE: Pennsylvania
CC	COUNTRY: USA
CC	ZIP: 19477
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Floppy disk
CC	COMPUTER: IBM PC compatible
CC	OPERATING SYSTEM: PC-DOS/MS-DOS
CC	SOFTWARE: PatentIn Release #1.0, Version #1.25
CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: US/08/115,680
CC	FILING DATE:
CC	CLASSIFICATION: 514
CC	PRIOR APPLICATION DATA:
CC	APPLICATION NUMBER: US 07/941,372
CC	FILING DATE: 02-SEP-1992
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: Bak, Mary E.
CC	REGISTRATION NUMBER: 31.215
CC	REFERENCE/DOCKET NUMBER: INDLAUSA
CC	TELECOMMUNICATION INFORMATION:
CC	TELEPHONE: (215) 540-9206
CC	TELEFAX: (215) 540-5818
CC	INFORMATION FOR SEQ ID NO: 2:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 199 amino acids
CC	TYPE: amino acid
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: protein
CC	SEQUENCE 199 AA; 21429 MW; 188641 CN;

\*\*\*\*\*  
W P E R E H (TM)  
\*\*\*\*\*

Release 3.1a John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

Mpsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:11:47 1999; MasPar time 6.52 Seconds  
234.409 Million cell updates/sec

Tabular output not generated.

Title: >US-08-938-548B-9  
Description: (1-28) from US08938548B.pep  
Perfect Score: 201  
Sequence: 1 RGPFGGLQRLRLQANGNHAAGILTM 28

Scoring table: PAM 150  
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrembl9  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 30.180; Variance 51.706; scale 0.584

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result	No.	Score	Match	Length	DB	ID	Pred. No.
1	201	100.0	130	11	05241	HYPOCRETIN (PREPRO-ORE	4.74e-26
2	201	100.0	130	11	05232	PREPRO-OREXIN.	4.74e-26
3	197	98.0	131	6	077668	PREPRO-OREXIN PRECURSO	3.52e-25
4	190	94.5	131	4	043612	PREPRO-OREXIN.	1.15e-23
5	76	37.8	1174	6	093168	TIGHT JUNCTION PROTEIN	2.96e-01
6	75	37.3	1277	11	035821	PAR INTERACTING PROTEI	4.31e-01
7	73	36.3	249	11	P97382	K+ CHANNEL BETA4 SUBUN	8.98e-01
8	71	35.3	145	10	048721	T9J22.21 PROTEIN.	1.85e+00
9	70	34.8	511	10	080740	T13D8.6 PROTEIN.	2.64e+00
10	69	34.3	331	13	Q91640	LEUCINE ZIPPER WITH BA	3.76e+00
11	69	34.3	389	11	O88539	MW DOMAIN BINDING PROT	3.76e+00
12	68	33.8	162	2	O85680	RNA POLYMERASE SIGMA-L	5.34e+00
13	68	33.8	481	3	042651	HYPOPHETICAL 55.5 KD P	5.34e+00
14	68	33.8	1015	10	064673	F22O13.25	5.34e+00
15	68	33.8	2115	4	O14980	NUMA PROTEIN.	5.34e+00
16	67	33.3	161	2	O85683	RNA POLYMERASE SIGMA-L	7.55e+00
17	67	33.3	178	2	O31955	YONC PROTEIN.	7.55e+00
18	67	33.3	178	9	O64066	HYPOPHETICAL 19.6 KD P	7.55e+00
19	67	33.3	413	11	O35141	MAXPL.	7.55e+00
20	67	33.3	462	2	P95644	RNA POLYMERASE SIGMA F	7.55e+00

21	67	33.3	510	2	059913	RNA POLYMERASE SIGMA F	7.55e+00
22	67	33.3	511	2	050539	MAJOR VEGETATIVE SIGMA	7.55e+00
23	67	33.3	514	2	P77951	RNA POLYMERASE SIGMA F	7.55e+00
24	67	33.3	525	2	Q59813	RNA POLYMERASE SIGMA F	7.55e+00
25	67	33.3	528	2	Q59814	RNA POLYMERASE SIGMA F	7.55e+00
26	67	33.3	541	2	O69851	HYPOTHETICAL 60.1 KD P	7.55e+00
27	67	33.3	568	5	Q27212	ARTICULIN P60.	7.55e+00
28	67	33.3	608	2	Q45998	DNA POLYMERASE III TAU	7.55e+00
29	66	32.8	130	2	O60245	HRPG.	1.07e+01
30	66	32.8	150	1	Q27642	DEOXYCYTIDINE-TRIPHOS	1.07e+01
31	66	32.8	253	2	O54224	UNKNOWN GENE.	1.07e+01
32	66	32.8	282	2	O87877	D-SUBUNIT OF BENZOYL-C	1.07e+01
33	66	32.8	304	5	Q22732	T24D5.1 PROTEIN.	1.07e+01
34	66	32.8	319	2	Q05804	RNA-DIRECTED DNA POLYM	1.07e+01
35	66	32.8	335	13	Q91654	THYROID HORMONE INDUCE	1.07e+01
36	66	32.8	365	5	Q09636	HYPOTHETICAL 41.3 KD P	1.07e+01
37	66	32.8	377	2	O50983	CARA.	1.07e+01
38	66	32.8	1465	5	O17909	H06O01.2 PROTEIN.	1.07e+01
39	66	32.8	2591	2	Q54959	PRISTINAMYCIN I SYNTHA	1.07e+01
40	65	32.3	213	2	O50648	TNA2, TNA1, PARTIAL AN	1.50e+01
41	65	32.3	443	2	O50205	TRANSCRIPTIONAL ACTIVA	1.50e+01
42	65	32.3	580	11	O60811	RETINOID X RECEPTOR IN	1.50e+01
43	65	32.3	1344	11	O35851	P160 MYB-BINDING PROTE	1.50e+01
44	65	32.3	1420	10	O81016	PUTATIVE ABC TRANSPORT	1.50e+01
45	65	32.3	2509	4	Q16702	FATTY ACID SYNTHASE (E	1.50e+01

ALIGNMENTS

RESULT 1  
ID O55241; PRELIMINARY; PRT; 130 AA.  
AC O55241;  
DT 01-JUN-1998 (TREMREL. 06, CREATED)  
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
DE HYPOCRETIN (PREPRO-OREXIN).  
GN HCRT.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98150861.  
RA SAKURAI T., AMEMIYA A., ISHII M., MATSUZAKI I., CHEMELLI R.M.,  
RA TANAKA H., WILLIAMS S.C., RICHARDSON J.A., KOZLOWSKI G.P., WILSON S.,  
RA ARCH J.R.S., BUCKINGHAM R.E., HAYNES A.C., CARR S.A., ANNAN R.S.,  
RA MCNULTY D.E., LIU W.-S., TERRETT J.A., ELSHOURBAGY N.A., BERGSMAN D.J.,  
RA YANAGISAWA M.;  
RT "Orexins and orexin receptors: a family of hypothalamic neuropeptides  
and G protein-coupled receptors that regulate feeding behavior.";  
RL CELL 92:573-585(1998).  
RN [12]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDUFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity.";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [13]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDUFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AF041242; G2897122; -  
DR EMBL; AF019566; G2895196; -  
DR MGD; MGI:1202306; HCRT.





Matches 8; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 280 LOEVLOSDSNHVPVSLT 296  
 ||:||||:||||:||||  
 Qy 11 LORLLQANGNHAAGILT 27

RESULT 14  
 ID 064673 PRELIMINARY; PRT: 1015 AA.  
 AC 064673;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
 DE F22013.25  
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
 OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA SHINN P., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y., SUN H.,  
 RA CONWAY A., CONWAY A., KURTZ D., OJI O., SHEN Y.K., TORIUMI M.,  
 RA VYSOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,  
 RA ECKER J.R.;  
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; AC003981; G3063483; -.  
 SQ SEQUENCE 1015 AA; 111751 MW; 82FA3C3F CRC32;

Query Match 33.8%; Score 68; DB 10; Length 1015;  
 Best Local Similarity 50.0%; Pred. No. 5.34e+00;  
 Matches 13; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Db 499 PTRGLRSSLRKRGNGGPTAATILT 524  
 | |||:: | || || || ||  
 Qy 2 FGPPGLQGLRQLQANGNHAAGILT 27

RESULT 15  
 ID Q14980 PRELIMINARY; PRT: 2115 AA.  
 AC Q14980;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE NUMA PROTEIN.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92176231.  
 RA YANG C.H., LAMBIE E.J., SNYDER M.;  
 RT "NUMA: an unusually long coiled-coil related protein in the mammalian  
 nucleus.";  
 RL J. CELL BIOL. 116:1303-1317(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94013066.  
 RA MAEKAWA T., KURIYAMA R.;  
 RT "Primary structure and microtubule-interacting domain of the SP-H  
 antigen: a mitotic MAP located at the spindle pole and characterized  
 as a homologous protein to NuMA.";  
 RL J. CELL SCI. 105:589-600(1993).  
 DR EMBL; 211583; G35119; -.  
 FT CONFLICT 124 124 Q -> P (IN REF. 2).  
 FT CONFLICT 1587 1587 Q -> H (IN REF. 2).  
 SQ SEQUENCE 2115 AA; 238273 MW; 81A36BA3 CRC32;

Query Match 33.8%; Score 68; DB 4; Length 2115;  
 Best Local Similarity 50.0%; Pred. No. 5.34e+00;  
 Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 761 RAGRKGLRLQLQGEAH 778  
 |:| ||::|||:|:|:

Qy 1 RGPPLGQLGRQLRLLQAN 18

Search completed: Fri Aug 20 21:12:40 1999  
 Job time : 53 secs.

\*\*\*\*\*

WIREH

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:11:16 1999; Maspar time 3.39 Seconds

Tabular output not generated. 233.546 Million cell updates/sec

Title: >US-08-938-548B-9  
Description: (1-28) from US08938548B.pep  
Perfect Score: 201  
Sequence: 1 RQGPGLQGRQLRLLQANGNHAAGILTM 28

Scoring table: PAM 150  
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 31.574; Variance 48.443; scale 0.652

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	71	35.3	187	1	SODC_HAEP	6.21e-01
2	71	35.3	187	1	SODC_HAEIN	6.21e-01
3	69	34.3	1992	1	TR12_HUMAN	1.33e+00
4	68	33.8	736	1	ALD_MOUSE	1.94e+00
5	68	33.8	879	1	YDHB_ECOLI	1.94e+00
6	67	33.3	79	1	NIFU_FRAAL	2.82e+00
7	67	33.3	442	1	HRDB_STRCO	2.82e+00
8	67	33.3	1157	1	PEX1_PICPA	2.82e+00
9	66	32.8	398	1	VE2_HPV63	4.07e+00
10	65	32.3	307	1	CC36_CAEEL	5.86e+00
11	65	32.3	368	1	COA1_POVLY	5.86e+00
12	65	32.3	799	1	AFSK_STRCO	5.86e+00
13	65	32.3	1692	1	CYAA_SCHPO	5.86e+00
14	64	31.8	199	1	IL11_MOUSE	8.40e+00
15	64	31.8	256	1	PRN3_HUMAN	8.40e+00
16	64	31.8	624	1	SIR_SYNP7	8.40e+00
17	64	31.8	1477	1	YOR1_YEAST	1.20e+01
18	63	31.3	255	1	YPEL_RHOU	1.20e+01
19	63	31.3	394	1	RT04_YEAST	1.20e+01
20	63	31.3	445	1	RF1M_HUMAN	1.20e+01
21	63	31.3	556	1	YEJH_ECOLI	1.20e+01
22	63	31.3	633	1	NODQ_RHSB	1.20e+01
23	63	31.3	702	1	YCBY_ECOLI	1.20e+01

24 63 31.3 878 1 ECR\_DROME ECDYSONE RECEPTOR (ECD 1.20e+01  
25 63 31.3 881 1 Y310\_HUMAN HYPOTHETICAL PROTEIN K 1.20e+01  
26 62 30.8 256 1 YREC\_SYNP2 HYPOTHETICAL 28.7 KD P 1.70e+01  
27 62 30.8 380 1 F812\_MOUSE FACTOR VIIII INTRON 22 1.70e+01  
28 62 30.8 400 1 ASSY\_SYNY3 ARGININOSUCCINATE SYNT 1.70e+01  
29 62 30.8 652 1 RPSD\_CAUCR RNA POLYMERASE SIGMA F 1.70e+01  
30 62 30.8 684 1 CALH\_HUMAN COLLAGEN ALPHA 1(XVIII) 1.70e+01  
31 62 30.8 1678 1 CA64\_HUMAN COLLAGEN ALPHA 6(IV) C 1.70e+01  
32 62 30.8 1690 1 CA44\_HUMAN COLLAGEN ALPHA 4(IV) C 1.70e+01  
33 62 30.8 1758 1 CA24\_CAEEL COLLAGEN ALPHA 2(IV) C 1.70e+01  
34 62 30.8 1763 1 CA24\_ASCSU PROCOLLAGEN ALPHA 2(IV 1.70e+01  
35 61 30.3 134 1 Y652\_METJA HYPOTHETICAL PROTEIN M 2.41e+01  
36 61 30.3 180 1 NEF\_HV2NZ NEGATIVE FACTOR (F-PRO 2.41e+01  
37 61 30.3 222 1 BASR\_SALTY TRANSCRIPTIONAL REGULA 2.41e+01  
38 61 30.3 240 1 YDGB\_ECOLI HYPOTHETICAL OXIDOREDU 2.41e+01  
39 61 30.3 248 1 PSPA\_HUMAN PULMONARY SURFACTANT-A 2.41e+01  
40 61 30.3 374 1 RGSX\_BOVIN RETINA SPECIFIC REGULA 2.41e+01  
41 61 30.3 409 1 ARR2\_HUMAN BETA-ARRESTIN 2 2.41e+01  
42 61 30.3 421 1 SAHL\_PYRHO ADENOSYLHOMOCYSTEINASE 2.41e+01  
43 61 30.3 516 1 LEGB\_GOSHI LEGUMIN B PRECURSOR (B 2.41e+01  
44 61 30.3 518 1 BAR2\_SCHCO PHEROMONE B ALPHA 2 RE 2.41e+01  
45 61 30.3 690 1 VTER\_EBV PROBABLE DNA PACKAGING 2.41e+01

#### ALIGNMENTS

RESULT 1  
ID SODC\_HAEP

AC P25842: STANDARD; PRT; 187 AA.  
DT 01-MAY-1992 (REL. 22, CREATED)  
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE SUPEROXIDE DISMUTASE [CU-ZN] PRECURSOR (EC 1.15.1.1).  
GN SODC.  
OS HAEMOPHILUS PARAINFLUENZAE.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;  
OC HAEMOPHILUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1391.  
RX MEDLINE; 92041655.  
RA KROLL J.S., LANGFORD P.R., LOYNDS B.M.;  
RT "Copper-zinc superoxide dismutase of Haemophilus influenzae and H.  
RT parainfluenzae.";  
RL J. BACTERIOL. 173:7449-7457(1991).  
CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.  
CC -!- FUNCTION: MAY CONFER SURVIVAL ADVANTAGE BY ACCELERATING  
CC DISMUTATION OF SUPEROXIDE OF ENVIRONMENTAL ORIGIN TO HYDROGEN  
CC PEROXIDE, DISRUPTIVE TO THE NORMAL MUCOCILIARY CLEARANCE PROCESS  
CC IN THE HOST.  
CC -!- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2).  
CC -!- SUBUNIT: HOMODIMER.  
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M84013; G148884; -  
CC PIR; B41654; B41654.  
CC PROSITE; PS00087; SOD\_CU\_ZN\_1; 1.  
CC PROSITE; PS00332; SOD\_CU\_ZN\_2; 1.  
CC PFW; PF00080; soccu; 1.  
CC HSP; P00446; IYAL.  
CC OXIDOREDUCTASE; COPPER; ZINC; PERIPLASMIC; SIGNAL.  
CC SIGNAL 1 23 POTENTIAL.  
CC CHAIN 24 187 SUPEROXIDE DISMUTASE [CU-ZN].

Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 289 KPSPGLOAKLASL 302  
QY 1 RPPGPGLOGLRL 14

## RESULT 4

ID ALD\_MOUSE STANDARD; PRT; 736 AA.  
AC P48410;  
DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE ADRENOLEUKODYSTROPHY PROTEIN HOMOLOG (ALDP).  
GN ALD OR ALDGH.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-FIBROTIC LIVER;  
RX MEDLINE; 95201443.  
RA SARDE C.O., THOMAS J., SADOULET H., GARNIER J.M., MANDEL J.L.;  
RT "CDNA sequence of Aldgh, the mouse homolog of the x-linked adrenoleukodystrophy gene";  
RL MAMM. GENOME 5:810-813(1994).  
CC -1- FUNCTION: PROBABLE TRANSPORTER. COULD BE INVOLVED IN THE IMPORT OF VLCFA-COA SYNTHETASE INTO THE PEROXISOMAL MEMBRANE.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL (PROBABLE).  
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). BELONGS TO THE MDR SUBFAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL; Z33637; G520955; -  
DR MGD; MGI:99672; ALDGH.  
DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
DR PFAM; PF00005; ABC\_tran; 1.  
KW ATP-BINDING; GLYCOPROTEIN; TRANSMEMBRANE; TRANSPORT; PEROXISOME.  
FT TRANSMEM 92 112 POTENTIAL.  
FT TRANSMEM 131 151 POTENTIAL.  
FT TRANSMEM 238 258 POTENTIAL.  
FT TRANSMEM 333 353 POTENTIAL.  
FT TRANSMEM 473 493 POTENTIAL.  
FT NP\_BIND 507 514 ATP (BY SIMILARITY).  
SQ SEQUENCE 736 AA; 81858 MW; D373B00E CRC32;

Query Match 33.8%; Score 68; DB 1; Length 736;  
Best Local Similarity 44.4%; Pred. No. 1.94e+00;  
Matches 12; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Db 704 AGTPKMGRIQLRQILGEAAAPVQL 730  
QY 2 PGPPGLOGLRLQANGNHAAGILTM 28

## RESULT 5

ID YDBH\_ECOLI STANDARD; PRT; 879 AA.  
AC P52645; P77502; P76855;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 96.8 KD PROTEIN IN LDHA-TYNA INTERGENIC REGION.  
GN YDBH.  
OS ESCHERICHIA COLI.  
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;

## OC ESCHERICHIA.

[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE; 97426617.  
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V., RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F., GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J., MAU B., SHAO Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL SCIENCE 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T., IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KANAI K., KASAI H., KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K., MORI H., MOTOMURA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y., YANO M.;  
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [3]  
RP SEQUENCE OF 1-331 FROM N.A.  
RC STRAIN-K12;  
RA BUNCH P.K., MAT-JAN F., LEE N.A., DEAYALA B.A., CLARK D.P.;  
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [4]  
RP IDENTIFICATION.  
RA RUDD K.E.;  
RL UNPUBLISHED OBSERVATIONS (MAR-1996).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL; AE000235; G1787646; -  
DR EMBL; D90776; G1742260; -  
DR EMBL; D90777; G1742264; -  
DR EMBL; U36928; -; NOT ANNOTATED\_CDS.  
KW ECOGENE; EGI1380; YDBH.  
KW HYPOTHETICAL PROTEIN.  
FT CONFLICT 36 36 I -> L (IN REF. 3).  
SQ SEQUENCE 879 AA; 96834 MW; 8BFD7CF3 CRC32;

Query Match 33.8%; Score 68; DB 1; Length 879;  
Best Local Similarity 45.8%; Pred. No. 1.94e+00;  
Matches 11; Conservative 6; Mismatches 6; Indels 1; Gaps 1;

Db 396 GVDGRQLQAHNELGDFVLHM 419  
QY 6 GLQGRQLRLQANGNHAAG-ILTM 28

RESULT 6  
ID NIFU\_FRAAL STANDARD; PRT; 79 AA.  
AC P46045;  
DT 01-NOV-1995 (REL. 32, CREATED)  
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE NIFU PROTEIN.  
GN NIFU.  
OS FRANKIA ALNI.  
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIADAE;  
OC ACTINOMYCETALES; FRANKINEAE; FRANKIACEAE; FRANKIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CP11;  
RX MEDLINE; 95369734.  
RA HARRIOTT O.T., HOSTED T.J., BENSON D.R.;

CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.  
 CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCCNNNGGT-3') PRESENT  
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER  
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION  
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS  
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION  
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA  
 CC REPLICATION.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- SUBUNIT: BINDS DNA AS A DIMER.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X70828; G312096; -  
 DR PFAM; PF00508; E2\_N; 1.  
 DR PFAM; PF00511; E2\_C; 1.  
 DR HSP; P17383; 1DHM.  
 KW EARLY PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING;  
 KW TRANS-ACTING FACTOR; DNA REPLICATION; REPRESSOR; NUCLEAR PROTEIN.  
 SQ SEQUENCE 398 AA; 45450 MW; C9BB0CE0 CRC32;  
 -----  
 Query Match 32.8%; Score 66; DB 1; Length 398;  
 Best Local Similarity 56.3%; Pred. No. 4.07e+00;  
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
 -----  
 Db 295 RSPPKGGQSRRLRLQ 310  
 QY 1 RGPGLGRLQLRLQ 16  
 I: | | | | | | | | | |  
 1: | | | | | | | | | |  
 -----  
 RESULT 10  
 ID CC36\_CAEEL STANDARD; PRT; 307 AA.  
 AC P34803;  
 DT 01-FEB-1994 (REL. 28, CREATED)  
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE CUTICLE COLLAGEN 36.  
 GN COL-36 OR C27H5.5.  
 OS CAENORHABDITIS ELEGANS.  
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTERA; RHABDITIDA;  
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE: 94131298.  
 RA LEVY A.D., KRAMER J.M.;  
 RT "Identification, sequence and expression patterns of the  
 RT Caenorhabditis elegans col-36 and col-40 collagen-encoding genes.";  
 RL GENE 137:281-285(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA PAULEY A.;  
 RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
 CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
 CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
 CC LINKS.  
 CC -!- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE  
 CC COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L15418; G289662; -  
 DR EMBL; U14635; G540271; ALT\_INIT.  
 DR WORMPEP; C27H5.5; CE06893.  
 KW CUTICLE; CONNECTIVE TISSUE; REPEAT; MULTIGENE FAMILY; COLLAGEN.  
 FT DOMAIN 89 105  
 FT DOMAIN 118 150  
 FT DOMAIN 167 187  
 FT DOMAIN 194 226  
 FT DOMAIN 231 237  
 FT DOMAIN 260 295  
 SQ SEQUENCE 307 AA; 30126 MW; 9346DA48 CRC32;  
 -----  
 Query Match 32.3%; Score 65; DB 1; Length 307;  
 Best Local Similarity 55.8%; Pred. No. 5.86e+00;  
 Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 -----  
 Db 214 PGPPGAPGPGRVIOVNG 231  
 QY 2 PGPPGLGRLQLRLQANG 19  
 I: | | | | | | | | | |  
 1: | | | | | | | | | |  
 -----  
 RESULT 11  
 ID COAL\_POVLY STANDARD; PRT; 368 AA.  
 AC P04010;  
 DT 23-OCT-1986 (REL. 02, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE COAT PROTEIN VP1.  
 OS LYMPHOTROPIC POLYOMAVIRUS.  
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; POLYOMAVIRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 86045879.  
 RA PAWLITA M., CLAD A., ZUR HAUSEN H.;  
 RT "Complete DNA sequence of lymphotropic papovavirus: prototype of a  
 RT new species of the polyomavirus genus.";  
 RL VIROLOGY 143:196-211(1985).  
 CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF FIVE COPIES OF VP1.  
 CC EACH OF WHICH IS COMPOSED OF 72 ICOSAEDRAL UNITS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; K02562; G333278; -  
 DR PIR; A03627; VVPIL.  
 DR PFAM; PF00718; Polyoma\_coat; 1.  
 DR HSP; P49302; 1SIE.  
 KW LATE PROTEIN; COAT PROTEIN.  
 SQ SEQUENCE 368 AA; 40211 MW; 4979282C CRC32;  
 -----  
 Query Match 32.3%; Score 65; DB 1; Length 368;  
 Best Local Similarity 32.0%; Pred. No. 5.86e+00;  
 Matches 8; Conservative 7; Mismatches 10; Indels 0; Gaps 0;  
 -----  
 Db 185 KPGNQGLDPKAKPLLDKGNYPVEY 209  
 QY 1 RGPGLGRLQLRLQANGNHAAGI 25  
 I: | | | | | | | | | |  
 1: | | | | | | | | | |  
 -----  
 RESULT 12  
 ID AFSK\_STRCO STANDARD; PRT; 799 AA.  
 AC P54741;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

Db 1539 PVOLQGRLERLIKS 1552  
 QY 4 PPGQLGRLLQQA 17

RESULT 14  
 ID IL11 MOUSE STANDARD; PRT; 199 AA.  
 AC P47873;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE INTERLEUKIN-11 PRECURSOR (IL-11).  
 GN IL11.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97070356.  
 RA MORRIS J.C., FINNERTY H., BENNETT F., TURNER K.J., WOOD C.R.;  
 RT "Molecular cloning and characterization of murine interleukin-11.";  
 RL EXP. HEMATOL. 24:1369-1376(1996).  
 CC -!- FUNCTION: THIS PROTEIN STIMULATES PLASMACYTOMA PROLIFERATION,  
 CC T-CELL-DEPENDENT DEVELOPMENT OF IMMUNOGLOBULIN-PRODUCING B  
 CC CELLS AND SYNERGIZES WITH IL-3 IN SUPPORTING MURINE  
 CC MEKAKARYOCYTE COLONY FORMATION (BY SIMILARITY).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U03421; G415634; -  
 DR MGD; MGI:107613; IL11.  
 KW CYTOKINE; GROWTH FACTOR; SIGNAL.  
 FT SIGNAL 1 21  
 FT CHAIN 22 199 INTERLEUKIN-11.  
 SQ SEQUENCE 199 AA; 21522 MW; 1CB30772 CRC32;

Query Match 31.8%; Score 64; DB 1; Length 199;  
 Best Local Similarity 53.3%; Pred. No. 8.40e+00;  
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 124 PELGALQARLERLLR 138  
 QY 2 PGPPGLQGRLLQ 16

RESULT 15  
 ID PRN3-HUMAN STANDARD; PRT; 256 AA.  
 AC P24158; P15637;  
 DT 01-APR-1990 (REL. 14, CREATED)  
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE MYELOBLASTIN PRECURSOR (EC 3.4.21.76) (LEUKOCYTE PROTEINASE 3) (PR-3)  
 DE (PR3) (AGP7) (WEGENER'S AUTOANTIGEN) (P29) (C-ANCA ANTIGEN).  
 GN PRN3 OR MBN.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92021028.  
 RA LABAYE C., MUSETTE P., CAYRE Y.E.;  
 RT "Wegener autoantigen and myeloblastin are encoded by a single mRNA.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:9253-9256(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA LAMERDIN J.E., MCCREADY P.M., SKOWRONSKI E., ADAMSON A.W.,

BURKHART-SCHULTZ K.J., GORDON L., KYLE A., RAMIREZ M., STILWAGEN S.,  
 PHAN H., VELASCO N., DO L., REGALA W., TERRY A., GARNES J.,  
 DANGANAN L., POUNDSTONE P., CHRISTENSEN M., GEORGESCU A., AVILA J.,  
 LIU S., ATTIX C., ANDREISE T., TRANKHEIM M., AMICO-KELLER G.,  
 COFIELD J., DUARTE S., LUCAS S., BRUCE R., THOMAS P., QUAN G.,  
 KRONMILLER B., ARELLANO A., MONTGOMERY M., OW D., NOLAN M., TRONG S.,  
 KOBAYASHI A., OLSEN A.S., CARRANO A.V.;  
 RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE OF 2-256 FROM N.A., AND SEQUENCE OF 48-71 AND 156-181.  
 RX MEDLINE; 91079774.  
 RA CAMPANELLI D., MELCHIOR M., FU Y., NAKATA M., SHUMAN H., NATHAN C.,  
 GABAY J.E.;  
 RT "Cloning of cDNA for proteinase 3: a serine protease, antibiotic, and  
 RT autoantigen from human neutrophils.";  
 RL J. EXP. MED. 172:1709-1715(1990).  
 RN [4]  
 RP SEQUENCE OF 22-256 FROM N.A.  
 RX MEDLINE; 92390417.  
 RA ZIMMER M., MEDCALF R.L., FINK T.M., MATTMANN C., LICHTER P.,  
 JENNE D.E.;  
 RT "Three human elastase-like genes coordinately expressed in the  
 RT myelomonocyte lineage are organized as a single genetic locus on  
 RT 19pter.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 89:8215-8219(1992).  
 RN [5]  
 RP SEQUENCE OF 42-256 FROM N.A.  
 RX MEDLINE; 90090622.  
 RA BORIES D., RAYNAL M.-C., SOLOMON D.H., DARZYNKIEWICZ Z., CAYRE Y.E.;  
 RT "Down-regulation of a serine protease, myeloblastin, causes growth  
 RT arrest and differentiation of promyelocytic leukemia cells.";  
 RL CELL 59:959-968(1989).  
 RN [6]  
 RP SEQUENCE OF 28-67 AND 228-244.  
 RX MEDLINE; 91236723.  
 RA RAO N.V., WEHNER N.G., MARSHALL B.C., GRAY W.R., GRAY B.H.,  
 HOIDAL J.R.;  
 RT "Characterization of proteinase-3 (PR-3), a neutrophil serine  
 RT proteinase. Structural and functional properties.";  
 RL J. BIOL. CHEM. 266:9540-9548(1991).  
 RN [7]  
 RP SEQUENCE OF 28-47.  
 RX MEDLINE; 89315847.  
 RA GABAY J.E., SCOTT R.W., CAMPANELLI D., GRIFFITH J., WILDE C.,  
 MARA M.N., SEGER M., NATHAN C.F.;  
 RT "Antibiotic proteins of human polymorphonuclear leukocytes.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 86:5610-5614(1989).  
 RN [8]  
 RP SEQUENCE OF 28-47 AND 196-219.  
 RX MEDLINE; 90130450.  
 RA WILDE C.G., SNABLE J.L., GRIFFITH J.E., SCOTT R.W.;  
 RT "Characterization of two azurophil granule proteases with active-site  
 RT homology to neutrophil elastase.";  
 RL J. BIOL. CHEM. 265:2038-2041(1990).  
 RN [9]  
 RP SEQUENCE OF 1-20 FROM N.A.  
 RX MEDLINE; 92390417.  
 RA ZIMMER M., MEDCALF R.L., FINK T.M., MATTMANN C., LICHTER P.,  
 JENNE D.E.;  
 RT "Three human elastase-like genes coordinately expressed in the  
 RT myelomonocyte lineage are organized as a single genetic locus on  
 RT 19pter.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 89:8215-8219(1992).  
 RN [10]  
 RP SEQUENCE OF 28-48, AND IDENTITY OF WEGENER'S AUTOANTIGEN WITH PR-3.  
 RX MEDLINE; 90332035.  
 RA JENNE D.E., TSCHOPP J., LUEDEMANN J., UTECHT B., GROSS W.L.;  
 RT "Wegener's autoantigen decoded.";  
 RL NATURE 346:520-520(1990).  
 RN [11]  
 RP IDENTITY OF WEGENER'S AUTOANTIGEN WITH PROTEINASE 3.  
 RX MEDLINE; 91055123.  
 RA GUPTA S.K., NILES J.L., MCCLUSKEY R.T., ARNAOUT M.A.;

(TM)

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	201	100.0	123	33	W61383	2.26e-11
2	201	100.0	130	33	W61382	2.26e-11
3	201	100.0	130	30	W50158	2.26e-11
4	201	100.0	130	30	W50157	2.26e-11
5	190	94.5	131	33	W61381	2.79e-10
6	68	33.8	2192	23	W21732	6.37e+01
7	68	33.8	2272	23	W21731	6.37e+01
8	66	32.8	10	4	R20235	9.27e+01
9	65	32.3	716	18	R39737	1.12e+02
10	65	32.3	1464	35	W64835	1.12e+02
11	65	32.3	2509	24	W32881	1.12e+02
12	64	31.8	256	9	R45403	1.34e+02
13	64	31.8	256	15	R85639	1.34e+02
14	64	31.8	549	16	R92050	1.34e+02
15	64	31.8	551	38	W83404	1.34e+02
16	64	31.8	1477	13	R67691	1.34e+02

CC antagonists, in therapy, to detect Ab and to isolate cognate  
CC receptors. Oligonucleotides based on H35 cDNA can be used to detect  
CC the hypocretin gene or its RNA transcript, and as antisense agents  
CC for inhibiting gene expression. H35 cDNA can also be used for  
CC recombinant protein production. The Ab can be used to detect or  
CC quantify hypocretin proteins and as a therapeutic inhibitor.  
SQ Sequence 130 AA;

Query Match 100.0%; Score 201; DB 30; Length 130;  
Best Local Similarity 100.0%; Pred. No. 2.26e-11;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dbb 69 rpppglqrlqlrqangnhaagiltm 96  
|||||  
Qy 1 RPPGPGLOGLRQLRLQANGNHAAGILTM 28

RESULT 5

ID W61381 standard; Protein: 131 AA.

AC W61381;  
DE Human HFGAN72 receptor protein.  
KW HFGAN72 receptor; eating disorders; renal disease; heart failure;  
KW sexual dysfunction; HIV; cancer; Parkinson's disease; urinary retention;  
KW osteoporosis; angina pectoris; myocardial infarction; psychotic disorder;  
KW neurological disorder.  
OS Homo sapiens.

FH Key Location/Qualifiers  
FT Region 34..56  
FT FT /note= "Ligand 72A"  
FT Region 70..97  
FT FT /note= "Ligand 72B"

EP-849361-A2.  
PD 24-JUN-1998.  
PF 17-DEC-1997; 310216.  
PR 26-SEP-1997; US-939093.  
PR 17-DEC-1996; US-033604.  
PR 19-MAR-1997; US-820519.  
PR 02-JUL-1997; US-887382.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA Bergsma DJ, Brooks DP, Gellai M, Wilson S, Yanagisawa M;  
PI WPI: 98-324672/29.  
DR N-PSDB: V28138.  
PT HFGAN72 receptor ligands - and corresponding DNA, agonists,  
PT antibodies, antagonists, etc.  
PS Claim 2; Fig 2; 35pp; English.

CC The HFGAN72 receptor protein contains two ligands whose antagonists can  
CC be used for treating obesity, diabetes, anorexia nervosa, bulimia  
CC cachexia, chronic renal failure, renal disease, congestive heart failure  
CC impaired glucose tolerance and sexual dysfunction. The agonist is  
CC useful for treating anorexia nervosa, bulimia and cachexia. The HFGAN72  
CC receptor ligand is useful for treating e.g. bacterial, fungal, protozoan  
CC and viral infections, particularly infections caused by HIV-1 or HTV-2,  
CC pain, cancers, anorexia nervosa, bulimia, cachexia, obesity, diabetes,  
CC asthma, Parkinson's disease, both acute and congestive heart failure,  
CC hypotension, hypertension, urinary retention, osteoporosis, angina  
CC pectoris, myocardial infarction, ulcers, asthma, allergies, benign  
CC prostatic hypertrophy, chronic renal failure, renal disease, impaired  
CC glucose tolerance, sexual dysfunction and psychotic and neurological  
CC disorders including anxiety, schizophrenia, manic depression, delirium,  
CC dementia, severe mental retardation and dyskinesias such as Huntington's  
CC disease or Gilles de la Tourette's syndrome.  
SQ Sequence 131 AA;

```

Query Match      94.5%; Score 190; DB 33; Length 131;
Best Local Similarity 92.9%; Pred. No. 2.79e-10;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 70 rsgppglqgrlqlrqasghaaagiltm 97
Qy 1 RPQPPGGLQGRLLQRLQANGNHAAGILTM 28

```

RESULT	6	
ID	W21732 standard; Protein; 2192 AA.	
AC	W21732;	
DT	01-OCT-1997 (first entry)	
DE	LexA/NumA fusion protein.	
KW	NIP-1; NIP-2; NuMA; nuclear mitotic apparatus; NuMA interacting protein;	
KW	cell division; proliferation; antibody; Ab; detection;	
KW	malignant cell growth.	
OS	Homo sapiens.	
FT	Key	Location/Qualifiers
FT	domain	1..87
FT	FT	/label= LexA_DNA_binding_domain
FT	peptide	88..94
FT	FT	/label= Polylinker
FT	protein	95..2192
FT	FT	/label= Residues_18-2116_of_NuMA
FT	region	285..1784
FT	FT	/label= Coiled_coil_region
PN	W09640917-A1..	
PD	19-DEC-1996.	
PF	07-JUN-1996; U09504.	
PR	07-JUN-1995; US-478408.	
PT	(UYUA ) UNIV YALE.	
PI	McPherson SMG, Snyder MP;	
DR	WPI: 97-077270/07.	
DR	N-PSDB: 777783.	
DR	New nucleic acid encoding nuclear mitotic appts. interacting	
PT	proteins - useful for modulating cell division and proliferation and	
PT	in diagnosis	
PS	Claim 15; Page 42-50; 78pp; English.	
CC	The sequences given in W21731-32 represent fusion proteins which contain	
CC	NuMA (nuclear mitotic apparatus). The fusion proteins were used in	
CC	the identification of NuMA interacting proteins (NIP's) (see also	
CC	W21729-30). Compounds which interfere with the interaction of NuMA	
CC	with a known NIP are used to modulate cell division and/or proliferation	
CC	Ab, raised conventionally using NIP-1 or -2 as immunogen, are used to	
CC	detect NIP (or their complexes) and to block their activity for	
CC	diagnostic or therapeutic use, e.g. to detect defective NuMA or NIP	
CC	which may be markers for aberrant (including malignant) cell growth	
CC	(which can also be detected by nucleic acid sequencing). Also where	
CC	malignancy is related to defects in NuMA or NIP, it can be treated by	
CC	administration of the appropriate functional protein.	
SQ	Sequence 2192 AA;	
SQ		

```

Query Match      33.8%   Score 68;   DB 23;   Length 2192;
Best Local Similarity 50.0%;   Pred. No. 6.37e+01;
Matches 9;   Conservative 6;   Mismatches 3;   Indels 0;   Gaps 0;

Db      838   ragrkglearqlqglesh 855

Qy      1   RPPPPGLQGRLLQAN 18
      | | | | | | | | | |

```

RESULT	7	
ID	W21731	standard; Protein; 2272 AA.
AC	W21731;	
DE	01-OCT-1997	(first entry)
DT	GAL4/HA/NumA	fusion protein.
KW	NIP-1;	NIP-2; NUMA; nuclear mitotic apparatus; NumA interacting protein;
KW	cell division;	proliferation; antibody; Ab; detection;
KW	malignant cell growth.	
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	domain	1..147
FT	peptide	/label= GAL4_DNA_binding_domain
FT	protein	148..174
FT	region	/label= Hemagglutinin_epitope
FT	region	175..2272
FT	region	/label= Residues_18-2116_of_NumA
FT	region	365..1864
FT	region	/label= Coiled_coil_region
PN	WO9640917-A1.	





\*\*\*\*\*

W P S R E H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:24:10 1999; MasPar time 1.85 Seconds

Tabular output not generated. 153.946 Million cell updates/sec

Title: >US-08-938-548B-12  
Description: (1-28) from US08938548B.pep  
Perfect Score: 201  
Sequence: 1 RGPPLGRLQLRLQANGNHAAGILTM 28

Scoring table: PAM 150  
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCT9\_COMB 4:backfiles1

Statistics: Mean 20.582; Variance 81.576; scale 0.252

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	67	33.3	442	2	US-08-363-Sequence 4, Applicatio	2.72e+01
2	67	33.3	442	2	US-08-363-Sequence 11, Applicati	2.72e+01
3	67	33.3	445	2	US-08-363-Sequence 5, Applicati	2.72e+01
4	67	33.3	445	2	US-08-363-Sequence 12, Applicati	2.72e+01
5	66	32.8	10	3	PCT-US91-0 Sequence 3, Applicatio	3.30e+01
6	65	32.3	716	3	PCT-US95-1 Sequence 4, Applicatio	4.00e+01
7	65	32.3	2509	1	US-08-469-Sequence 10, Applicatio	4.00e+01
8	63	31.3	1841	2	US-08-804-Sequence 6, Applicatio	5.87e+01
9	62	30.8	1694	1	US-08-494-Sequence 2, Applicatio	7.10e+01
10	61	30.3	381	2	US-08-858-Sequence 3, Applicatio	8.57e+01
11	61	30.3	381	2	US-08-858-Sequence 1, Applicatio	8.57e+01
12	60	29.9	199	3	PCT-US93-0 Sequence 2, Applicatio	1.03e+02
13	60	29.9	199	1	US-07-949-Sequence 4, Applicatio	1.03e+02
14	60	29.9	199	1	US-08-115-Sequence 2, Applicatio	1.03e+02
15	60	29.9	199	1	US-08-792-Sequence 8, Applicatio	1.03e+02
16	60	29.9	199	2	US-08-814-Sequence 4, Applicatio	1.03e+02
17	60	29.9	199	1	US-08-017-Sequence 4, Applicatio	1.03e+02
18	60	29.9	199	1	US-07-941-Sequence 2, Applicatio	1.03e+02
19	60	29.9	296	1	US-07-921-Sequence 14, Applicati	1.03e+02
20	60	29.9	296	3	PCT-US93-0 Sequence 4, Applicatio	1.03e+02
21	60	29.9	296	3	PCT-US94-1 Sequence 14, Applicatio	1.03e+02
22	60	29.9	296	1	US-08-115-Sequence 4, Applicatio	1.03e+02
23	60	29.9	296	1	US-07-941-Sequence 4, Applicatio	1.03e+02

24	60	29.9	296	1	US-08-165-Sequence 14, Applicati	1.03e+02
25	60	29.9	296	1	US-07-745-Sequence 14, Applicati	1.03e+02
26	60	29.9	300	2	US-08-794-Sequence 2, Applicatio	1.03e+02
27	60	29.9	405	1	US-07-688-Sequence 14, Applicati	1.03e+02
28	60	29.9	405	3	PCT-US91-0 Sequence 14, Applicati	1.03e+02
29	60	29.9	528	2	US-08-363-Sequence 14, Applicati	1.03e+02
30	60	29.9	530	2	US-08-363-Sequence 8, Applicatio	1.03e+02
31	60	29.9	530	2	US-08-363-Sequence 10, Applicati	1.03e+02
32	60	29.9	530	2	US-08-363-Sequence 3, Applicatio	1.03e+02
33	60	29.9	530	2	US-08-363-Sequence 2, Applicatio	1.03e+02
34	60	29.9	530	2	US-08-363-Sequence 9, Applicatio	1.03e+02
35	59	29.4	116	1	US-08-687-Sequence 3, Applicatio	1.25e+02
36	59	29.4	116	2	US-08-816-Sequence 3, Applicatio	1.25e+02
37	59	29.4	176	3	PCT-US96-1 Sequence 6, Applicatio	1.25e+02
38	59	29.4	176	2	US-08-659-Sequence 6, Applicatio	1.25e+02
39	59	29.4	430	2	US-08-318-Sequence 16, Applicatio	1.25e+02
40	59	29.4	588	2	US-08-620-Sequence 2, Applicatio	1.25e+02
41	59	29.4	3724	2	US-08-804-Sequence 10, Applicati	1.25e+02
42	58	28.9	199	4	5215895-4 Patent No. 5215895.	1.50e+02
43	58	28.9	226	4	5498600-2 Patent No. 5498600.	1.50e+02
44	58	28.9	241	4	5175255-8 Patent.No. 5175255.	1.50e+02
45	58	28.9	1367	1	US-07-946-Sequence 4, Applicatio	1.50e+02

#### ALIGNMENTS

RESULT 1  
ID US-08-363-255-4 STANDARD; PRT; 442 AA.  
XX  
AC xxxxxx  
XX  
DT  
DE  
DE Sequence 4, Application US/08363255  
XX  
XX  
CC Sequence 4, Application US/08363255  
CC Patent No. 5783386  
CC GENERAL INFORMATION:  
CC APPLICANT: JACOBS, JR., WILLIAM R.  
CC APPLICANT: BLOOM, BARRY R.  
CC APPLICANT: COLLINS, DESMOND M.  
CC APPLICANT: de LISLE, GEOFFREY W.  
CC APPLICANT: PASCOPELLA, LISA  
CC APPLICANT: KAWAKAMI, RIKU P.  
CC TITLE OF INVENTION: MYCOBACTERIA VIRULENCE FACTORS AND A  
CC NUMBER OF SEQUENCES: 14  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: MORRISON & FOERSTER  
CC STREET: 755 Page Mill Road  
CC CITY: Palo Alto  
CC STATE: California  
CC COUNTRY: USA  
CC ZIP: 94304-1018  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/363/255  
CC FILING DATE: 23-DEC-1994  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: MONROY, GLADYS H.  
CC REGISTRATION NUMBER: 32,430  
CC REFERENCE/DOCKET NUMBER: 25237-20002.22  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 813-5600  
CC TELEFAX: (415) 494-0792  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:

CC	TELEPHONE: 202-639-7700
CC	TELEFAX: 202-639-7890
CC	TELEX:
CC	INFORMATION FOR SEQ ID NO: 10:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 2509 amino acids
CC	TYPE: amino acid
CC	STRANDEDNESS: single
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: protein
CC	HYPOTHETICAL: NO
CC	ANTI-SENSE: NO
CC	FRAGMENT TYPE: internal
CC	ORIGINAL SOURCE:
SQ	SEQUENCE 2509 AA; 273089 MW; 32011381 CN;
Query Match 32.3%; Score 65; DB 1; Length 2509;	
Best Local Similarity 29.6%; Pred. NO. 4.00e+01;	
Matches 8; Conservative 10; Mismatches 9; Indels 0; Gaps 0;	
Db	411 PPAPAPHATPLRLLRASGRTPAEVKL 437
Qy	2 PGPGLQGRQLRLLQANGNHAAGILTM 28
: : :       :   : : :	
RESULT 8	
ID	US-08-804-227C-6 STANDARD; PRT; 1841 AA.
XX	
AC	xxxxxx
XX	
DT	
XX	
DE	
XX	Sequence 6, Application US/08804227C
CC	
CC	Sequence 6, Application US/08804227C
CC	Patent No. 5876991
CC	GENERAL INFORMATION:
CC	APPLICANT: DeHoff, Bradley S.
CC	APPLICANT: Kuhstoss, Stuart A.
CC	APPLICANT: Rostek, Paul R., Jr.
CC	APPLICANT: Sutton, Kimberly L.
CC	TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
CC	NUMBER OF SEQUENCES: 15
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: THOMAS G. PLANT 1501
CC	STREET: LILLY CORPORATE CENTER
CC	CITY: INDIANAPOLIS
CC	STATE: IN
CC	COUNTRY: USA
CC	ZIP: 46285
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Floppy disk
CC	COMPUTER: IBM Compatible
CC	OPERATING SYSTEM: MS-DOS
CC	SOFTWARE: ASCII(DOS) Text only
CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: US/08/804,227C
CC	FILING DATE: February 21, 1997
CC	CLASSIFICATION: 435
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: Plant, Thomas, G.
CC	REGISTRATION NUMBER: 35,784
CC	REFERENCE/DOCKET NUMBER: X-8231
CC	TELECOMMUNICATION INFORMATION:
CC	TELEPHONE: 317-276-2459
CC	INFORMATION FOR SEQ ID NO: 6:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 1841 amino acids
CC	TYPE: amino acid
CC	TOPOLOGY: unknown
CC	MOLECULE TYPE: protein
SQ	SEQUENCE 1841 AA; 191190 MW; 14819276 CN;

```
Query Match      31.3%; Score 63; DB 2; Length 1841;  
Best Local Similarity 33.3%; Pred. No. 5.87e+01;  
Matches        8; Conservative    8; Mismatches   8; Indels     0; Gaps       0;
```

D**b** 1513 RDTPAALAAHAEELATARDGPG 1536  
  
QY 1 RGPFGIQLRLORLLQANGHAAG 24  
| : | : | : | : | : | : | : |  
| : | : | : | : | : | : | :

R**E**SULT 9  
ID US-08-494-168-2 STANDARD; PRT; 1694 AA.  
XX AC xxxxxx  
XX DT  
XX XX  
DE XX  
XX XX

Sequence 2, Application US/08494168

CC Sequence 2, Application US/08494168  
CC Patent No. 5731192  
CC GENERAL INFORMATION:  
CC APPLICANT: Readers, Stephen T.  
CC APPLICANT: Zhou, Jing  
CC TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method  
CC TITLE OF INVENTION: of Detecting Collagen Deficiency  
CC NUMBER OF SEQUENCES: 10  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Foley & Lardner  
CC STREET: 3000 K Street, N.W., Suite 500  
CC CITY: Washington, D.C.  
CC COUNTRY: USA  
CC ZIP: 20007-5109

COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patentin Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/494,168  
CC FILING DATE:  
CC CLASSIFICATION: 435

PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/112,465  
CC FILING DATE: 27-AUG-1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Saxe, Bernhard D.  
CC REGISTRATION NUMBER: 28,665

REFERENCE/DOCKET NUMBER: 40397/104/BABR

TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (202)672-5300  
CC TELEFAX: (202)672-5399  
CC TELEX: 904136

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1694 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein

SQ SEQUENCE 1694 AA; 165283 MW; 14956750 CN;

Query Match 30.8%; Score 62; DB 1; Length 1694;  
Best Local Similarity 88.9%; Pred. No. 7.10e+01;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

D**b** 1028 RGPFPGLPG 1036  
|||||||  
|

QY 1 RPGPSGLQG 9

R**E**SULT 10  
ID US-08-858-052-3 STANDARD; PRT; 381 AA.  
XX AC xxxxxx

CC CLASSIFICATION: 530  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Meinert, M. C.  
CC REGISTRATION NUMBER: 31,544  
CC REFERENCE/DOCKET NUMBER: 5174BPCT  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 199 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 199 AA; 21429 MW; 188641 CN;

Query Match 29.9%; Score 60; DB 1; Length 199;  
Best Local Similarity 70.0%; Pred.No. 1.03e+02;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 129 LQARLDRLRL 138  
QY 7 LQGRLORLQ 16  
||:|||||:

RESULT 14  
ID US-08-115-680-2 STANDARD; PRT; 199 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE  
XX  
Sequence 2, Application US/08115680  
Sequence 2, Application US/08115680  
Patent No. 5437863  
GENERAL INFORMATION:  
CC APPLICANT: Williams, David A.  
CC APPLICANT: Clark, Steven C.  
CC TITLE OF INVENTION: Method of Treating Cell Damage or  
CC TITLE OF INVENTION: Depletion  
CC NUMBER OF SEQUENCES: 4  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Howson and Howson  
CC STREET: Spring House Corporate Cntr, P.O. Box 457  
CC CITY: Spring House  
CC STATE: Pennsylvania  
CC COUNTRY: USA  
CC ZIP: 19477  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/115,680  
CC FILING DATE:  
CC CLASSIFICATION: 514  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 07/941,372  
CC FILING DATE: 02-SEP-1992  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Bak, Mary E.  
CC REGISTRATION NUMBER: 31,215  
CC REFERENCE/DOCKET NUMBER: IND1AUSA  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (215) 540-9206  
CC TELEFAX: (215) 540-5818  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 199 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
SQ SEQUENCE 199 AA; 21429 MW; 188641 CN;

Query Match 29.9%; Score 60; DB 1; Length 199;  
Best Local Similarity 70.0%; Pred.No. 1.03e+02;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 129 LQARLDRLRL 138  
QY 7 LQGRLORLQ 16  
||:|||||:

RESULT 15  
ID US-08-792-019B-8 STANDARD; PRT; 199 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE  
XX  
Sequence 8, Application US/08792019B  
Sequence 8, Application US/08792019B  
Patent No. 5741772  
GENERAL INFORMATION:  
CC APPLICANT: CHANG, MING-SHI  
CC TITLE OF INVENTION: THE NEUROTROPHIC FACTOR NNT-1  
CC NUMBER OF SEQUENCES: 16  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: AMGEN INC.  
CC STREET: 1840 DEHAVILLAND DRIVE  
CC CITY: THOUSAND OAKS  
CC STATE: CA  
CC COUNTRY: USA  
CC ZIP: 91320  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/792,019B  
CC FILING DATE: 03-FEB-1997  
CC CLASSIFICATION: 514  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: COOK, ROBERT R.  
CC REGISTRATION NUMBER: 31,602  
CC REFERENCE/DOCKET NUMBER: A-442  
CC INFORMATION FOR SEQ ID NO: 8:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 199 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC FEATURE:  
CC NAME/KEY: Protein  
CC LOCATION: 1..178  
CC NAME/KEY: Region  
CC LOCATION: -21..0  
SQ SEQUENCE 199 AA; 21429 MW; 188641 CN;

Query Match 29.9%; Score 60; DB 1; Length 199;  
Best Local Similarity 70.0%; Pred.No. 1.03e+02;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 129 LQARLDRLRL 138  
QY 7 LQGRLORLQ 16  
||:|||||:

Search completed: Fri Aug 20 21:24:17 1999  
Job time : 7 secs.

\*\*\*\*\*

WQSPFH (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:21:48 1999; MasPar time 4.63 Seconds

Tabular output not generated. 242.105 Million cell updates/sec

Title: >US-08-938-548B-12

Description: (1-28) from US08938548B.pep

Perfect Score: 201

Sequence: 1 RPPGPGQLGRLLQANGNHAGILTM 28

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 30.698; Variance 52.943; scale 0.580

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	71	35.3	145	2	T00987	hypothetical protein
2	71	35.3	187	2	B41654	superoxide dismutase
3	71	35.3	187	2	A41654	superoxide dismutase
4	70	34.8	511	2	T02669	hypothetical protein
5	68	33.8	736	2	S47044	malpde protein - mouse
6	68	33.8	879	2	H64888	membrane protein ydbH
7	68	33.8	1015	2	T00730	hypothetical protein
8	68	33.8	2115	2	S23647	hUMA protein - human
9	67	33.3	178	2	E69313	hypothetical protein
10	67	33.3	442	2	S11712	transcription initiat
11	67	33.3	510	2	S41307	transcription initiat
12	67	33.3	525	2	JN0443	transcription initiat
13	67	33.3	528	2	JN0445	transcription initiat
14	67	33.3	1157	2	A55152	PAS1 protein - yeast
15	66	32.8	150	2	D69081	deoxyuridine 5-tripho
16	66	32.8	253	2	S49183	hypothetical protein
17	66	32.8	319	1	S19248	RNA-directed DNA poly
18	65	32.3	1420	2	T02644	probable ABC transpor
19	65	32.3	1692	2	A33988	adenylate cyclase (EC
20	65	32.3	2309	2	G01880	fatty-acid synthase (
21	64	31.8	239	2	S25618	hypothetical protein
22	64	31.8	256	1	PRH03	proteinase 3 (EC 3.4.
23	64	31.8	263	2	S51155	hypothetical protein

24	64	31.8	439	2	S61858	hrpE protein - Pseudo	1.92e+01
25	64	31.8	535	2	A46101	protein-tyrosine-phos	1.92e+01
26	64	31.8	548	2	B46101	protein-tyrosine-phos	1.92e+01
27	64	31.8	624	1	RDYCS7	sulfite reductase (fe	1.92e+01
28	64	31.8	1348	2	S27812	probable epidermal ce	1.92e+01
29	64	31.8	1348	2	A43917	probable epidermal ce	1.92e+01
30	64	31.8	1477	2	S64616	TOR1 protein - yeast	1.92e+01
31	63	31.3	255	2	S12255	hypothetical protein	1.92e+01
32	63	31.3	394	2	A42115	ribosomal protein S2,	2.67e+01
33	63	31.3	556	2	B64939	hypothetical protein	2.67e+01
34	63	31.3	702	2	C64835	hypothetical protein	2.67e+01
35	63	31.3	770	2	S77523	hypothetical protein	2.67e+01
36	63	31.3	812	2	S31521	collagen COL1 - fres	2.67e+01
37	63	31.3	878	2	A41055	ecdysone receptor - f	2.67e+01
38	62	30.8	380	2	A42832	factor VIII-associate	3.68e+01
39	62	30.8	466	2	S61292	transcription initiat	3.68e+01
40	62	30.8	652	2	I40676	transcription initiat	3.68e+01
41	62	30.8	684	2	A53019	collagen alpha 1(XVII	3.68e+01
42	62	30.8	767	2	T00360	hypothetical protein	3.68e+01
43	62	30.8	1288	3	JE0363	mitogen-activated pro	3.68e+01
44	62	30.8	1690	1	CGHUIB	collagen alpha 4(IV)	3.68e+01
45	62	30.8	1763	2	S16366	collagen alpha 2(IV)	3.68e+01

#### ALIGNMENTS

RESULT 1  
ENTRY T00987 #type complete  
TITLE hypothetical protein T9J22.21 - Arabidopsis thaliana  
ORGANISM #formal\_name Arabidopsis thaliana #common\_name mouse-ear  
cress  
DATE 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change  
12-Feb-1999  
ACCESSIONS T00987  
REFERENCE Z14153  
#authors Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;  
Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.;  
Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter,  
J.C.

#submission submitted to the EMBL Data Library, April 1998  
#description Arabidopsis thaliana chromosome II BAC T9J22 genomic  
sequence.  
#accession T00987

##status preliminary; translated from GB/EMBL/DBJ  
##molecule\_type DNA  
##residues 1-145 #label ROU  
##cross-references EMBL:AC002505; NID:g2739359; PID:g2739379

#### GENETICS

#map\_position II  
#introns 17/1; 49/3; 78/2; 123/2  
#note T9J22.21

SUMMARY #length 145 #molecular-weight 15355 #checksum 3045

Query Match 35.3%; Score 71; DB 2; Length 145;  
Best Local Similarity 39.1%; Pred. No. 1.77e+00;  
Matches 9; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

Db 116 PERPGLGQVSTMEALGAHADS 138

| |||::: |||

QY 2 PGPPGLQRLRLQANGNHAG 24

#### RESULT 2

ENTRY B41654 #type complete  
TITLE superoxide dismutase (EC 1.15.1.1) (Cu-Zn) precursor -  
Haemophilus parainfluenzae  
ORGANISM #formal\_name Haemophilus parainfluenzae  
DATE 12-Jun-1992 #sequence\_revision 12-Jun-1992 #text\_change  
05-Mar-1999

#### ACCESSIONS

REFERENCE B41654

#authors Kroll, J.S.; Langford, P.R.; Loynds, B.M.

#journal J. Bacteriol. (1991) 173:7449-7457

```

6
RESULT      6
ENTRY       H64888      #type complete
TITLE       membrane protein ydbH - Escherichia coli
ORGANISM    12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
DATE        13-Nov-1998

ACCESSIONS  H64888
REFERENCE    A64720
#authors    Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
            Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
            Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
            Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
            Y.
#journal    Science (1997) 277:1453-1462
#title      The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession  H64888
#status     nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues   1-879 #label BLAT
#cross-references GB:AF000235; GB:U00096; NID:g1787643; PID:g1787646;
            WGP:b1381
#experimental_source strain K-12, substrain MG1655

GENETICS    ydbH
#gene       ydbH
FEATURE     8-24
SUMMARY     #domain transmembrane #status predicted #label TM01
            #length 879 #molecular-weight 96834 #checksum 5911

Query Match      33.8%; Score 68; DB 2; Length 879;
Best Local Similarity 45.8%; Pred. No. 5.02e+00;
Matches 11; Conservative 6; Mismatches 6; Indels 1; Gaps 1;

Db 396 GVDGRQLQAHLNGLGDFVLHM 419
      1:|||||:||||:|:|:|
QY 6 GLOGRLQLQANGNHAAG-ILTM 28
      1:|||||:||||:|:|:|

7
RESULT      7
ENTRY       T00730      #type complete
TITLE       hypothetical protein F22013.25 - Arabidopsis thaliana
ORGANISM    12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
DATE        12-Feb-1999

ACCESSIONS  T00730
REFERENCE    Z14200
#authors    Shinn, P.; Buehler, E.; Dewar, K.; Peng, J.; Kim, C.; Li, Y.;
            Sun, H.; Conway, A.; Conway, A.; Kurtz, D.; Oji, O.; Shen,
            Y.K.; Toriumi, M.; Vysotskaia, V.; Yu, G.; Davis, R.W.;
            Federspiel, N.A.; Theologis, A.; Ecker, J.R.
            submitted to the EMBL Data Library, April 1998
#submission Genomic sequence for Arabidopsis thaliana BAC F22013.
#accession  T00730
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues    1-1015 #label SHI
#cross-references EMBL:AC003981; NID:g3063438; PID:g3063463

GENETICS    I
#map_position 286/3; 508/3; 684/3; 722/3; 749/3; 785/2; 813/3
#introns     #length 1015 #molecular-weight 111751 #checksum 6299

SUMMARY     33.8%; Score 68; DB 2; Length 1015;
Query Match      33.8%; Score 68; DB 2; Length 1015;
Best Local Similarity 50.0%; Pred. No. 5.02e+00;
Matches 13; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Db 499 PTPRGLRSSLRKRNNGNPTAATILT 524
      1:|||||:|:|:|
QY 2 PGPPGLQRLQLQANGNHAAGILT 27
      1:|||||:|:|:|

8
RESULT      8

```

```

6
RESULT      6
ENTRY       H64888      #type complete
TITLE       membrane protein ydbH - Escherichia coli
ORGANISM    12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
DATE        13-Nov-1998

ACCESSIONS  H64888
REFERENCE    A64720
#authors    Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
            Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
            Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
            Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
            Y.
#journal    Science (1997) 277:1453-1462
#title      The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession  H64888
#status     nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues   1-879 #label BLAT
#cross-references GB:AF000235; GB:U00096; NID:g1787643; PID:g1787646;
            WGP:b1381
#experimental_source strain K-12, substrain MG1655

GENETICS    ydbH
#gene       ydbH
FEATURE     8-24
SUMMARY     #domain transmembrane #status predicted #label TM01
            #length 879 #molecular-weight 96834 #checksum 5911

Query Match      33.8%; Score 68; DB 2; Length 879;
Best Local Similarity 45.8%; Pred. No. 5.02e+00;
Matches 11; Conservative 6; Mismatches 6; Indels 1; Gaps 1;

Db 396 GVDGRQAILQAHENELGDFVLHM 419
      1:|||||:||||:|:|:|
QY 6 GLOGRLQRLQANGHAAAG-ILTM 28
      1:|||||:||||:|:|:|

7
RESULT      7
ENTRY       T00730      #type complete
TITLE       hypothetical protein F22013.25 - Arabidopsis thaliana
ORGANISM    12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
DATE        12-Feb-1999

ACCESSIONS  T00730
REFERENCE    214200
#authors    Shinn, P.; Buehler, E.; Dewar, K.; Peng, J.; Kim, C.; Li, Y.;
            Sun, H.; Conway, A.; Conway, A.; Kurtz, D.; Oji, O.; Shen,
            Y.K.; Toriumi, M.; Vysotskaia, V.; Yu, G.; Davis, R.W.;
            Federspiel, N.A.; Theologis, A.; Ecker, J.R.
            submitted to the EMBL Data Library, April 1998
#submission Genomic sequence for Arabidopsis thaliana BAC F22013.
#accession  T00730
#status     preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues    1-1015 #label SHI
#cross-references EMBL:AC003981; NID:g3063438; PID:g3063463

GENETICS    I
#map_position 286/3; 508/3; 684/3; 722/3; 749/3; 785/2; 813/3
#introns     #length 1015 #molecular-weight 111751 #checksum 6299

SUMMARY     33.8%; Score 68; DB 2; Length 1015;
            Best Local Similarity 50.0%; Pred. No. 5.02e+00;
            Matches 13; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Query Match      33.8%; Score 68; DB 2; Length 1015;
Best Local Similarity 50.0%; Pred. No. 5.02e+00;
Matches 13; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Db 499 PTPRGLRSSLRKRNNGNPAAITLT 524
      1:|||||:|:|:|
QY 2 PGPPGLQRLQRLQANGNHAAGILT 27
      1:|||||:|:|:|

8
RESULT      8

```

```

##status      preliminary
##molecule_type  DNA
##residues    1-1157 ##label HEY
##cross-references EMBL:Z36987; NID:g537419; PID:g537420
GENETICS
  $gene
CLASSIFICATION  #superfamily FtsH/SEC18/CDC48-type ATP-binding domain
                 homology
                 ATP; P-loop; peroxisome biogenesis
KEYWORDS
523-530        #region nucleotide-binding motif A (P-loop)\
813-1020       #domain FtsH/SEC18/CDC48-type ATP-binding domain
                 homology #label VAPN
840-847       #region nucleotide-binding motif A (P-loop)
SUMMARY        #length 1157 #molecular-weight 126983 #checksum 2540

Query Match      33.3%; Score 67; DB 2; Length 1157;
Best Local Similarity 60.0%; Pred. No. 7.06e+00;
Matches          9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 1072 KLEHLYOQGNHAEQ 1086
      ::::|:|:|:|:|
QY  10 RLQRLQANGNHAAG 24

RESULT 15
ENTRY   D69081      #type complete
TITLE   deoxyuridine 5-triphosphate nucleotidohydrolase related
        protein - Methanobacterium thermoautotrophicum (strain
        Delta H)
ORGANISM #formal_name Methanobacterium thermoautotrophicum
DATE      05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
        05-Jun-1998
ACCESSIONS D69081
REFERENCE  A69000
          $authors
            Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
            Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
            Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
            Lumm, W.; Pothier, B.; Olu, D.; Spadafora, R.; Vicaire, R.;
            Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso,
            A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
            McDougall, S.; Shimer, G.; Goyal, A.; Pietrokowski, S.;
            Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
            J.; Reeve, J.N.
          $journal
            J. Bacteriol. (1997) 179:7135-7155
          $title
            Complete genome sequence of Methanobacterium
            thermoautotrophicum Delta H: functional analysis and
            comparative genomics.
          $cross-references EMBL:Z36987; NID:g537419; PID:g537420
          $accession
            D69081
          $status
            preliminary; nucleic acid sequence not shown;
            translation not shown
          $molecule_type
            DNA
          $residues
            1-150 ##label MTH
          $cross-references GB:AF000920; GB:AF000666; NID:g5622729; PID:g2622730
          $experimental_source strain Delta H
GENETICS
  $gene
SUMMARY        MTH1605
                 #length 150 #molecular-weight 16996 #checksum 5013

Query Match      32.8%; Score 66; DB 2; Length 150;
Best Local Similarity 50.0%; Pred. No. 9.90e+00;
Matches          9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 102 GDGFRGTQLQFLNHGE 119
      |:::|:|:|:|:|
QY  3 GPPGLQRLQRLQANGN 20

Search completed: Fri Aug 20 21:22:21 1999
Job time : 33 secs.

```



\*\*\*\*\*

W P E R E H

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:22:38 1999; MasPar time 3.26 Seconds  
Tabular output not generated. 242.897 Million cell updates/sec

Title: >US-08-938-548B-12  
Description: (1-28) from US08938548B.pep  
Perfect Score: 201  
Sequence: 1 RCPPLQGLRQLRLLQANGNHAAGILTM 28

Scoring table: PAM 150  
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:swissprot

Statistics: Mean 31.574; Variance 48.443; scale 0.652

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	71	35.3	187	1 SDC_HAEP	SUPEROXIDE DISMUTASE [	6.21e-01
2	71	35.3	187	1 SDC_HAEP	SUPEROXIDE DISMUTASE [	6.21e-01
3	69	34.3	1992	1 TR12_HUMAN	THYROID RECEPTOR INTER	1.33e+00
4	68	33.8	736	1 ADL_MOUSE	ADRENOLKODYSTROPHY P	1.94e+00
5	68	33.8	879	1 YDHB_ECOLI	HYPOTHETICAL 96.8 KD P	1.94e+00
6	67	33.3	79	1 NIFU_FRAAL	NIFU PROTEIN.	2.82e+00
7	67	33.3	442	1 HKB_STRCO	RNA POLYMERASE PRINCIP	2.82e+00
8	67	33.3	1157	1 PEK1_PICPA	PEROXISOME BIOSYNTHESI	2.82e+00
9	66	32.8	398	1 V22_HRV63	REGULATORY PROTEIN E2.	4.07e+00
10	65	32.3	307	1 CC36_CAEEL	CUTICLE COLLAGEN 36.	5.86e+00
11	65	32.3	368	1 COAL_POVLY	COAT PROTEIN VP1.	5.86e+00
12	65	32.3	799	1 AFSK_STRCO	SERINE/THREONINE PROTE	5.86e+00
13	65	32.3	1692	1 CYAA_SCHPO	ADENYLATE CYCLASE (EC	5.86e+00
14	64	31.8	199	1 ILL1_MOUSE	INTERLEUKIN-11 PRECURS	8.40e+00
15	64	31.8	256	1 PRN3_HUMAN	MYELOBLASTIN PRECURSOR	8.40e+00
16	64	31.8	624	1 SIR_SYNP7	SULFITE REDUCTASE (FER	8.40e+00
17	64	31.8	1477	1 YOR1_YEAST	OLIGOMYCIN RESISTANCE	8.40e+00
18	63	31.3	255	1 YPE1_RHOU	HYPOTHETICAL 28 KD PRO	1.20e+01
19	63	31.3	394	1 RT04_YEAST	MITOCHONDRIAL 40S RIBO	1.20e+01
20	63	31.3	445	1 RF1M_HUMAN	MITOCHONDRIAL PEPTIDE	1.20e+01
21	63	31.3	556	1 YEAJ_ECOLI	HYPOTHETICAL 63.2 KD P	1.20e+01
22	63	31.3	633	1 NODQ_RHIB	PROBABLE SULFATE ADENY	1.20e+01
23	63	31.3	702	1 YCBY_ECOLI	HYPOTHETICAL 78.9 KD P	1.20e+01

24	63	31.3	878	1 ECR_DROME	ECDYSONE RECEPTOR (ECD	1.20e+01
25	63	31.3	881	1 Y31Q_HUMAN	HYPOTHETICAL PROTEIN K	1.20e+01
26	62	30.8	256	1 YREC_SINP2	HYPOTHETICAL 28.7 KD P	1.70e+01
27	62	30.8	380	1 FB12_MOUSE	FACTOR VIII INTRON 22	1.70e+01
28	62	30.8	400	1 ASSY_SYNY3	ARGININOSUCCINATE SYNT	1.70e+01
29	62	30.8	652	1 RPSD_CAUCR	RNA POLYMERASE SIGMA F	1.70e+01
30	62	30.8	684	1 CALH_HUMAN	COLLAGEN ALPHA 1(XVII)	1.70e+01
31	62	30.8	1678	1 CA64_HUMAN	COLLAGEN ALPHA 6(IV) C	1.70e+01
32	62	30.8	1690	1 CA44_HUMAN	COLLAGEN ALPHA 4(IV) C	1.70e+01
33	62	30.8	1758	1 CA24_CAEEL	COLLAGEN ALPHA 2(IV) C	1.70e+01
34	62	30.8	1763	1 CA24_ASCS	PROCOLLAGEN ALPHA 2(IV	1.70e+01
35	61	30.3	134	1 Y652_METJA	HYPOTHETICAL PROTEIN M	2.41e+01
36	61	30.3	180	1 NEF_HV2NZ	NEGATIVE FACTOR (F-PRO	2.41e+01
37	61	30.3	222	1 BASR_SALTY	TRANSCRIPTIONAL REGULA	2.41e+01
38	61	30.3	240	1 YDGB_ECOLI	HYPOTHETICAL OXIDOREDU	2.41e+01
39	61	30.3	248	1 PSPA_HUMAN	PULMONARY SURFACTANT-A	2.41e+01
40	61	30.3	374	1 RGSX_BOVIN	RETINA SPECIFIC REGULA	2.41e+01
41	61	30.3	409	1 ARR2_HUMAN	BETA-ARRESTIN 2.	2.41e+01
42	61	30.3	421	1 SAKH_PYRHO	ADENOSYLHOMOCYSTEINASE	2.41e+01
43	61	30.3	516	1 LEGB_GOSHI	LEGUMIN B PRECURSOR (B	2.41e+01
44	61	30.3	518	1 BAR2_SCHCO	PHEROMONE B ALPHA 2 RE	2.41e+01
45	61	30.3	690	1 VTER_EBV	PROBABLE DNA PACKAGING	2.41e+01

#### ALIGNMENTS

RESULT	1			
ID	SODC_HAEP	STANDARD;	PRT;	187 AA.
AC	P25842;			
DT	01-MAY-1992 (REL. 22, CREATED)			
DT	01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)			
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)			
DE	SUPEROXIDE DISMUTASE [CU-ZN] PRECURSOR (EC 1.15.1.1).			
GN	SODC.			
OS	HAEMOPHILUS PARAINFLUENZAE.			
OC	BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;			
CC	HAEMOPHILUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=1391;			
RX	MEDLINE; 92041655.			
RA	KROLL J.S., LANGFORD P.R., LOYNDS B.M.;			
RT	"Copper-zinc superoxide dismutase of Haemophilus influenzae and H.			
RT	parainfluenzae.";			
RL	J. BACTERIOL. 173:7449-7457(1991).			
CC	!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE			
CC	CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.			
CC	!- FUNCTION: MAY CONFER SURVIVAL ADVANTAGE BY ACCELERATING			
CC	DISMUTATION OF SUPEROXIDE OF ENVIRONMENTAL ORIGIN TO HYDROGEN			
CC	PEROXIDE, DISRUPTIVE TO THE NORMAL MUCOCILIARY CLEARANCE PROCESS			
CC	IN THE HOST.			
CC	!- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2).			
CC	!- SUBUNIT: HOMODIMER.			
CC	!- SUBCELLULAR LOCATION: PERIPLASMIC (POTENTIAL).			
CC	!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; M84013; G148884; -			
DR	PIR; B41654; B41654.			
DR	PROSITE; PS00087; SOD_CU_ZN_1; 1.			
DR	PROSITE; PS00332; SOD_CU_ZN_2; 1.			
DR	PFAM; PF00080; sodcu; 1.			
DR	HSP; P00446; IYAI.			
DR	OXIDOREDUCTASE; COPPER; ZINC; PERIPLASMIC; SIGNAL.			
KW	SIGNAL	1	23	POTENTIAL.
FT	CHAIN	24	187	SUPEROXIDE DISMUTASE [CU-ZN]

Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 Db 289 KPSPGLOAKLASL 302  
 QY 1 RPPPGLOQLRL 14

RESULT 4  
 ID ALD\_MOUSE STANDARD; PRT; 736 AA.

AC P48410;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE ADRENOLEUKODYSTROPHY PROTEIN HOMOLOG (ALDP).  
 GN ALD OR ALDGH.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=FIBROTIC LIVER;  
 RX MEDLINE; 95201443.  
 RA SARDE C.O., THOMAS J., SADOULET H., GARNIER J.M., MANDEL J.L.;  
 RT "cDNA sequence of Aldgh, the mouse homolog of the x-linked  
 adrenoleukodystrophy gene";  
 RL MAMM. GENOME 5:810-813(1994).  
 CC -1- FUNCTION: PROBABLE TRANSPORTER. COULD BE INVOLVED IN THE  
 CC IMPORT OF VLCFA-COA SYNTHETASE INTO THE PEROXISOMAL MEMBRANE.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL  
 CC (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
 CC (ABC TRANSPORTERS). BELONGS TO THE MDR SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; Z33637; G520955;  
 DR MG0; MGI:199672; ALDGH.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
 DR PFAM; PF00005; ABC\_tran; 1.  
 KW ATP-BINDING; GLYCOPROTEIN; TRANSMEMBRANE; TRANSPORT; PEROXISOME.  
 FT TRANSMEM 92 112 POTENTIAL.  
 FT TRANSMEM 131 151 POTENTIAL.  
 FT TRANSMEM 238 258 POTENTIAL.  
 FT TRANSMEM 333 353 POTENTIAL.  
 FT TRANSMEM 473 493 POTENTIAL.  
 FT NP\_BIND 507 514 ATP (BY SIMILARITY).  
 SQ SEQUENCE 736 AA; 81958 MW; D373800E CRC32;

Query Match 33.8%; Score 68; DB 1; Length 736;  
 Best Local Similarity 44.4%; Pred. No. 1.94e+00;  
 Matches 12; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Db 704 AGIPKMQRLQELQILGEAAPVPL 730  
 QY 2 GPFPGLQRLQRLQANGNHAAGILTM 28

RESULT 5

ID YDBH\_ECOLI STANDARD; PRT; 879 AA.  
 AC P52645; P77502; P76855;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 96.8 KD PROTEIN IN LDHA-TYNA INTERGENIC REGION.  
 GN YDBH.  
 OS ESCHERICHIA COLI.  
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;

OC ESCHERICHIA.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE; 97426617.  
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,  
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,  
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GORDEN M.A., ROSE D.J.,  
 RA MAU B., SHAO Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL SCIENCE 277:1453-1474(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,  
 RA IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KANAI K., KASAI H.,  
 RA KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,  
 RA MASUDA S., MIKI T., MIZOBUCHI K., MORI H., MOTOMURA K., NAKAMURA Y.,  
 RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,  
 RA TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y., YANO M.;  
 RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [3]  
 RP SEQUENCE OF 1-331 FROM N.A.  
 RC STRAIN=K12;  
 RA BUNCH P.K., MAT-JAN F., LEE N.A., DEAYALA B.A., CLARK D.P.;  
 RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [4]  
 RP IDENTIFICATION.  
 RA RUDD K.E.;  
 RL UNPUBLISHED OBSERVATIONS (MAR-1996).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AE000235; G1787646;  
 DR EMBL; D90776; G1742260;  
 DR EMBL; D90777; G1742264;  
 DR EMBL; U36928; -; NOT\_ANNOTATED\_CDS.  
 DR ECOGENE; EG13180; YDBH.  
 KW HYPOTHETICAL PROTEIN.  
 FT CONFLICT 36 36 I -> L (IN REF. 3).  
 SQ SEQUENCE 879 AA; 96834 MW; 8BFD7CF3 CRC32;

Query Match 33.8%; Score 68; DB 1; Length 879;  
 Best Local Similarity 45.8%; Pred. No. 1.94e+00;  
 Matches 11; Conservative 6; Mismatches 6; Indels 1; Gaps 1;

Db 396 GVDGRLQAILQAHENELGDFVLHM 419  
 QY 6 GLQGRQLRLQANGNHAAG-ILTM 28

RESULT 6

ID NIFU\_FRAAL STANDARD; PRT; 79 AA.  
 AC P46045;  
 DT 01-NOV-1995 (REL. 32, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE NIFU PROTEIN.  
 GN NIFU.

OS FRANKIA ALNI.  
 OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;  
 OC ACTINOMYCETALES; FRANKINEAE; FRANKIACEAE; FRANKIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CP11;  
 RX MEDLINE; 95369734.  
 RA HARRIOTT O.T., HOSTED T.J., BENSON D.R.;

CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.  
CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACNNNNNGGT-3') PRESENT  
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER  
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION  
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS  
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION  
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA  
CC REPLICATION.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- SUBUNIT: BINDS DNA AS A DIMER.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X70828; G312096; -  
CC PFAM; PF00508; E2\_N; 1.  
CC PFAM; PF00511; E2\_C; 1.  
CC HSSP; P17383; 1DHM.  
CC  
CC EARLY PROTEIN: TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING;  
CC TRANS-ACTING FACTOR; DNA REPLICATION; REPRESSOR; NUCLEAR PROTEIN.  
CC  
CC SEQUENCE 398 AA; 45450 MW; C9BBOCE0 CRC32;  
CC  
CC  
CC Query Match 32.8%; Score 66; DB 1; Length 398;  
CC Best Local Similarity 56.3%; Pred. No. 4.07e+00;  
CC Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
CC  
CC Db 295 RSPKPGGSRRLRLIQ 310  
CC I: | | | | | | | | | |  
CC Y 1 RPPGPGGRLQLRLQ 16  
CC  
CC  
CC RESULT 10  
CC ID CC36\_CAEEL STANDARD; PRT; 307 AA.  
CC AC P34803;  
CC DT 01-FEB-1994 (REL. 28, CREATED)  
CC DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)  
CC DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
CC DE CUTICLE COLLAGEN 36.  
CC GN COL-36 OR C27H5.5.  
CC OS CAENORHABDITIS ELEGANS.  
CC OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA;  
CC OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RC STRAIN-BRISTOL N2;  
CC RX MEDLINE; 94131298.  
CC RA LEVY A.D., KRAMER J.M.;  
CC RT "Identification, sequence and expression patterns of the  
CC RT Caenorhabditis elegans col-36 and col-40 collagen-encoding genes.";  
CC RL GENE 137:281-285(1993).  
CC RN [2]  
CC RP SEQUENCE FROM N.A.  
CC RC STRAIN-BRISTOL N2;  
CC RA PAULEY A.;  
CC RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
CC LINKS.  
CC -!- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE  
CC COLLAGENS.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L15418; G289662; -  
CC EMBL; U14635; G540271; ALT\_INIT.  
CC DR WORMPEP; C27H5.5; CE06893.  
CC KW CUTICLE; CONNECTIVE TISSUE; REPEAT; MULTIGENE FAMILY; COLLAGEN.  
CC FT DOMAIN 89 105  
CC FT DOMAIN 118 150  
CC FT DOMAIN 167 187  
CC FT DOMAIN 194 226  
CC FT DOMAIN 231 257  
CC FT DOMAIN 260 295  
CC FT DOMAIN 307 AA; 30126 MW; 9346DA48 CRC32;  
CC SQ  
CC  
CC Query Match 32.3%; Score 65; DB 1; Length 307;  
CC Best Local Similarity 55.6%; Pred. No. 5.86e+00;  
CC Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
CC  
CC Db 214 PGPPGPGQGRVQVNG 231  
CC I: | | | | | | | | | |  
CC Y 2 PGPPGGLQRLQLRLQANG 19  
CC  
CC  
CC RESULT 11  
CC ID COAL\_POVLY STANDARD; PRT; 368 AA.  
CC AC P04010;  
CC DT 23-OCT-1986 (REL. 02, CREATED)  
CC DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
CC DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
CC DE COAT PROTEIN VP1.  
CC OS LYMPHOTROPIC POLYOMAVIRUS.  
CC OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; POLYOMAVIRUS.  
CC RN [1]  
CC RP SEQUENCE FROM N.A.  
CC RX MEDLINE; 86045879.  
CC RA PAWLITA M., CLAD A., ZUR HAUSEN H.;  
CC RT "Complete DNA sequence of lymphotropic papovavirus: prototype of a  
CC RT new species of the polyomavirus genus";  
CC RL VIROLOGY 143:196-211(1985).  
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 72 ICOSAEDRAL UNITS,  
CC EACH OF WHICH IS COMPOSED OF FIVE COPIES OF VP1.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; K02562; G333278; -  
CC DR PIR; A03627; VVPFIL.  
CC DR PFAM; PF00718; Polyoma\_coat; 1.  
CC DR HSSP; P49302; 1STE.  
CC KW LATE PROTEIN; COAT PROTEIN.  
CC SQ SEQUENCE 368 AA; 40211 MW; 4979282C CRC32;  
CC  
CC Query Match 32.3%; Score 65; DB 1; Length 368;  
CC Best Local Similarity 32.0%; Pred. No. 5.86e+00;  
CC Matches 8; Conservative 7; Mismatches 10; Indels 0; Gaps 0;  
CC  
CC Db 185 KPGNQGLDKPAKPLLDKGNYPVEV 209  
CC I: | | | | | | | | | |  
CC Y 1 RPPGPGGLQRLQLRLQANGNHAAGI 25  
CC  
CC  
CC RESULT 12  
CC ID AFSK\_STRCO STANDARD; PRT; 799 AA.  
CC AC P54741;  
CC DT 01-OCT-1996 (REL. 34, CREATED)  
CC DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

Db 1539 PVOLOGLRLRLIKS 1552  
 QY 4 PPGLOGLRLQLQA 17

RESULT 14  
 ID IL1L\_MOUSE STANDARD; PRT; 199 AA.  
 AC P47873;  
 DT 01-FEB-1996 (REL. 33, CREATED)  
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE INTERLEUKIN-11 PRECURSOR (IL-11).  
 GN IL1L.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIUROGATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97070356.  
 RA MORRIS J.C., FINNERTY H., BENNET F., TURNER K.J., WOOD C.R.;  
 RT "Molecular cloning and characterization of murine interleukin-11.";  
 RL EXP. HEMATOL. 24:1369-1376(1996).  
 CC -1- FUNCTION: THIS PROTEIN STIMULATES PLASMACYTOMA PROLIFERATION,  
 CC T-CELL-DEPENDENT DEVELOPMENT OF IMMUNOGLOBULIN-PRODUCING B  
 CC CELLS AND SYNERGIZES WITH IL-3 IN SUPPORTING MURINE  
 CC MEGAKARYOCYTE COLONY FORMATION (BY SIMILARITY).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U03421; G415654; .  
 DR MGD; MGI:107613; IL1L.  
 DR CYTOKINE; GROWTH FACTOR; SIGNAL.  
 KW SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 199 INTERLEUKIN-11.  
 SQ SEQUENCE 199 AA; 21522 MW; 1CB30772 CRC32;

Query Match 31.8%; Score 64; DB 1; Length 199;  
 Best Local Similarity 53.3%; Pred. No. 8.40e+00;  
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 124 PELCALQRLRLRL 138  
 QY 2 PGPGLOGLRLQLQ 16

RESULT 15  
 ID PRN3\_HUMAN STANDARD; PRT; 256 AA.  
 AC P24158; P15637;  
 DT 01-APR-1990 (REL. 14, CREATED)  
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE MYELOBLASTIN PRECURSOR (EC 3.4.21.76) (LEUKOCYTE PROTEINASE 3) (PR-3)  
 DE (PR3) (AGP7) (WEGENER'S AUTOANTIGEN) (P29) (C-ANCA ANTIGEN).  
 GN PRN3 OR MBN.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92021028.  
 RA LABBAYE C., MUSETTE P., CAYRE Y.E.;  
 RT "Wegener autoantigen and myeloblastin are encoded by a single mRNA.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:9253-9256(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA LAMERDIN J.E., MCCREADY P.M., SKOWRONSKI E., ADAMSON A.W.,

RA BURKHART-SCHULTZ K.J., GORDON L., KYLE A., RAMIREZ M., STILWAGEN S.,  
 RA PHAN H., VELASCO N., DO L., REGALA W., TERRY A., GARNES J.,  
 RA DANGANAN L., POUNDSTONE P., CHRISTENSEN M., GEORGESCU A., AVILA J.,  
 RA LIU S., ATTIX C., ANDREISE T., FRANKHEIM M., AMICO-KELLER G.,  
 RA COFIELD J., DUARTE S., LUCAS S., BRUCE R., THOMAS P., QUAN G.,  
 RA KRONMILLER B., ARELLANO A., MONTGOMERY M., OW D., NOLAN M., TRONG S.,  
 RA KOBAYASHI A., OLSEN A.S., CARRANO A.V.;  
 RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE OF 2-256 FROM N.A., AND SEQUENCE OF 48-71 AND 156-181.  
 RX MEDLINE; 91079774.  
 RA CAMPANELLI D., MELCHIOR M., FU Y., NAKATA M., SHUMAN H., NATHAN C.,  
 RA GABAY J.E.;  
 RT "Cloning of cDNA for proteinase 3: a serine protease, antibiotic, and  
 RT autoantigen from human neutrophils.";  
 RL J. EXP. MED. 172:1709-1715(1990).  
 RN [4]  
 RP SEQUENCE OF 2-256 FROM N.A.  
 RX MEDLINE; 92390417.  
 RA ZIMMER M., MEDCALF R.L., FINK T.M., MATTMANN C., LICHTER P.,  
 RA JENNE D.E.;  
 RT "Three human elastase-like genes coordinately expressed in the  
 RT myelomonocyte lineage are organized as a single genetic locus on  
 RT 19pter.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 89:8215-8219(1992).  
 RN [5]  
 RP SEQUENCE OF 42-256 FROM N.A.  
 RX MEDLINE; 90090622.  
 RA BORTES D., RAYNAL M.-C., SOLOMON D.H., DARZYNKIEWICZ Z., CAYRE Y.E.;  
 RT "Down-regulation of a serine protease, myeloblastin, causes growth  
 RT arrest and differentiation of promyelocytic leukemia cells.";  
 RL CELL 59:959-968(1989).  
 RN [6]  
 RP SEQUENCE OF 28-67 AND 228-244.  
 RX MEDLINE; 91236723.  
 RA RAO N.V., WEHNER N.G., MARSHALL B.C., GRAY W.R., GRAY B.H.,  
 RA HOIDAL J.R.;  
 RT "Characterization of proteinase-3 (PR-3), a neutrophil serine  
 RT proteinase. Structural and functional properties.";  
 RL J. BIOL. CHEM. 266:9540-9548(1991).  
 RN [7]  
 RP SEQUENCE OF 28-47.  
 RX MEDLINE; 89315847.  
 RA GABAY J.E., SCOTT R.W., CAMPANELLI D., GRIFFITH J., WILDE C.,  
 RA MARRA M.N., SEEGER M., NATHAN C.F.;  
 RT "Antibiotic proteins of human polymorphonuclear leukocytes.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 86:5610-5614(1989).  
 RN [8]  
 RP SEQUENCE OF 28-47 AND 196-219.  
 RX MEDLINE; 90130450.  
 RA WILDE C.G., SNABLE J.L., GRIFFITH J.E., SCOTT R.W.;  
 RT "Characterization of two azurophil granule proteases with active-site  
 RT homology to neutrophil elastase.";  
 RL J. BIOL. CHEM. 265:2038-2041(1990).  
 RN [9]  
 RP SEQUENCE OF 1-20 FROM N.A.  
 RX MEDLINE; 92390417.  
 RA ZIMMER M., MEDCALF R.L., FINK T.M., MATTMANN C., LICHTER P.,  
 RA JENNE D.E.;  
 RT "Three human elastase-like genes coordinately expressed in the  
 RT myelomonocyte lineage are organized as a single genetic locus on  
 RT 19pter.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 89:8215-8219(1992).  
 RN [10]  
 RP SEQUENCE OF 28-48, AND IDENTITY OF WEGENER'S AUTOANTIGEN WITH PR-3.  
 RX MEDLINE; 90332035.  
 RA JENNE D.E., TSCHOPP J., LUEDEMANN J., UTECHT B., GROSS W.L.;  
 RT "Wegener's autoantigen decoded.";  
 RL NATURE 346:520-520(1990).  
 RN [11]  
 RP IDENTITY OF WEGENER'S AUTOANTIGEN WITH PROTEINASE 3.  
 RX MEDLINE; 91055123.  
 RA GUPTA S.K., NILES J.L., MCCLUSKEY R.T., ARNAOUT M.A.;

\*\*\*\*\*

W P E R L H (TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:23:12 1999; MasPar time 6.52 Seconds  
234.276 Million cell updates/sec

Tabular output not generated.

Title: >US-08-938-548B-12

Description: (1-28) from US08938548B.pep

Perfect Score: 201

Sequence: 1 RGPFPGLQRLQRLQANGNHAAGILTM 28

Scoring table: PAM 150

Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: sptrembl9

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 30.180; Variance 51.706; scale 0.584

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	201	100.0	130	11	HYPOCRETIN (PREPRO-ORE	4.74e-26
2	201	100.0	130	11	PREPRO-OREXIN.	4.74e-26
3	197	98.0	131	6	PREPRO-OREXIN PRECURSO	3.52e-25
4	190	94.5	131	4	PREPRO-OREXIN.	1.15e-23
5	76	37.8	1174	6	TIGHT JUNCTION PROTEIN	2.98e-01
6	75	37.3	1277	11	PAR INTERACTING PROTEI	4.31e-01
7	73	36.3	249	11	K+ CHANNEL BETA4 SUBUN	8.98e-01
8	71	35.3	145	10	T9J22.21 PROTEIN.	1.85e+00
9	70	34.8	511	10	T13D8.6 PROTEIN.	2.64e+00
10	69	34.3	331	13	LEUCINE ZIPPER WITH BA	3.78e+00
11	69	34.3	389	11	WW DOMAIN BINDING PROT	3.76e+00
12	68	33.8	162	2	RNA POLYMERASE SIGMA-L	5.34e+00
13	68	33.8	481	3	HYPOTHETICAL 55.5 KD P	5.34e+00
14	68	33.8	1015	10	F22013.25.	5.34e+00
15	68	33.8	2115	4	NUMA PROTEIN.	5.34e+00
16	67	33.3	161	2	RNA POLYMERASE SIGMA-L	7.55e+00
17	67	33.3	178	2	YONC PROTEIN.	7.55e+00
18	67	33.3	178	9	HYPOTHETICAL 19.6 KD P	7.55e+00
19	67	33.3	413	11	MAXP1.	7.55e+00
20	67	33.3	462	2	RNA POLYMERASE SIGMA F	7.55e+00

21	67	33.3	510	2	059913	RNA POLYMERASE SIGMA F	7.55e+00
22	67	33.3	511	2	050539	MAJOR VEGETATIVE SIGMA	7.55e+00
23	67	33.3	514	2	P77951	RNA POLYMERASE SIGMA F	7.55e+00
24	67	33.3	525	2	Q59813	RNA POLYMERASE SIGMA F	7.55e+00
25	67	33.3	528	2	Q59814	RNA POLYMERASE SIGMA F	7.55e+00
26	67	33.3	541	2	Q69851	HYPOTHETICAL 60.1 KD P	7.55e+00
27	67	33.3	568	5	Q27212	ARTICULIN P60.	7.55e+00
28	67	33.3	608	2	Q45998	DNA POLYMERASE III TAU	7.55e+00
29	66	32.8	130	2	Q60245	HRPG.	1.07e+01
30	66	32.8	150	1	Q27642	DEOXYCYTIDINE-TRIPHOS	1.07e+01
31	66	32.8	253	2	Q54224	UNKNOWN GENE.	1.07e+01
32	66	32.8	282	2	Q87877	D-SUBUNIT OF BENZOYL-C	1.07e+01
33	66	32.8	304	5	Q22732	T24D5.1 PROTEIN.	1.07e+01
34	66	32.8	319	2	Q05804	RNA-DIRECTED DNA POLYM	1.07e+01
35	66	32.8	335	13	Q91654	THYROID HORMONE INDUCE	1.07e+01
36	66	32.8	365	5	Q09636	HYPOTHETICAL 41.3 KD P	1.07e+01
37	66	32.8	377	2	Q50983	CARA.	1.07e+01
38	66	32.8	1465	5	Q17909	H06O01.2 PROTEIN.	1.07e+01
39	66	32.8	2591	2	Q54959	PRISTINAMYCIN I SYNTHA	1.07e+01
40	65	32.3	213	2	Q50648	TNA2, TNA1, PARTIAL AN	1.50e+01
41	65	32.3	443	2	Q50205	TRANSCRIPTIONAL ACTIVA	1.50e+01
42	65	32.3	580	11	Q60811	RETINOID X RECEPTOR IN	1.50e+01
43	65	32.3	1344	11	Q35851	P160 MYB-BINDING PROTE	1.50e+01
44	65	32.3	1420	10	Q81016	POTATIVE ABC TRANSPORT	1.50e+01
45	65	32.3	2509	4	Q16702	FATTY ACID SYNTHASE (E	1.50e+01

## ALIGNMENTS

RESULT 1  
ID O55241; PRELIMINARY; PRT; 130 AA.  
AC O55241;  
DT 01-JUN-1998 (TREMREL. 06, CREATED)  
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
DE HYPOCRETIN (PREPRO-OREXIN).  
GN HCRT.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98150861.  
RA SAKURAI T., AMEMIYA A., ISHII M., MATSUZAKI I., CHEMELLI R.M.,  
RA TANAKA H., WILLIAMS S.C., RICHARDSON J.A., KOZLOWSKI G.P., WILSON S.,  
RA ARCH J.R.S., BUCKINGHAM R.E., HAYNES A.C., CARR S.A., ANNAN R.S.,  
RA MCNULTY D.E., LIU W.-S., TERRETT J.A., ELSHOUBAGY N.A., BERGSMAN D.J.,  
RA YANAGISAWA M.;  
RT "Orexins and orexin receptors: a family of hypothalamic neuropeptides  
RT and G protein-coupled receptors that regulate feeding behavior.";  
RL CELL 92:573-585(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN-C57BL/6J;  
RA DE LECHE L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
RT activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX STRAIN-C57BL/6J;  
RA DE LECHE L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
RL EMBL; AF041242; G2897122;  
DR EMBL; AF019566; G2895196;  
DR MGD; MGI:1202306; HCRT.

RT tumor suppressor protein.";  
 RL J. CELL BIOL. 124:949-961(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 96421547.  
 RA BEATCH M., JESAITIS L.A., GALLIN W., GOODENOUGH D.A., STEVENSON B.R.;  
 RT "The tight junction protein ZO-2 contains three PDZ  
 RT (PSD-95/Discs-Large/ZO-1) domains and an alternatively spliced  
 RT region.";  
 RL J. BIOL. CHEM. 271:25723-25726(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA GOODENOUGH D.A.;  
 RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA BEATCH M.;  
 RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; L27152; G1536570;  
 DR PFAM; PF00595; PDZ; 3;  
 DR PFAM; PF00625; Guanylate\_kin; 1.  
 SQ SEQUENCE 1174 AA; 132085 MW; 2FA16B83 CRC32;  
  
 Query Match 37.8%; Score 76; DB 6; Length 1174;  
 Best Local Similarity 32.1%; Pred. No. 2.98e-01;  
 Matches 9; Conservative 12; Mismatches 7; Indels 0; Gaps 0;  
  
 Db 268 RSPPELRGRPDHAGQPSDRPIGVLLM 295  
 QY 1 RPPPGLOGRLQLQANGHAAIGILTM 28  
  
 RESULT 6  
 ID O35821 PRELIMINARY; PRT; 1277 AA.  
 AC O35821;  
 DT 01-JAN-1998 (TREMREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
 DE PAR INTERACTING PROTEIN.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 91084854.  
 RA WUARIN J., SCHIBLER U.;  
 RT "Expression of the liver-enriched transcriptional activator protein  
 RT DBP follows a stringent circadian rhythm.";  
 RL CELL 63:1257-1266(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA COMTE P.A., OSSIPPOV V., SCHIBLER U.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; U83590; G2253211;  
 SQ SEQUENCE 1277 AA; 144674 MW; 89721F79 CRC32;  
  
 Query Match 37.3%; Score 75; DB 11; Length 1277;  
 Best Local Similarity 29.6%; Pred. No. 4.31e-01;  
 Matches 8; Conservative 14; Mismatches 5; Indels 0; Gaps 0;  
  
 Db 826 PGAEALHAQVERFVQAGNOADASVAL 852  
 QY 2 PGPPGLOGRLQLQANGHAAIGILTM 28  
  
 RESULT 7  
 ID P97382 PRELIMINARY; PRT; 249 AA.  
 AC P97382;  
 DT 01-MAY-1997 (TREMREL. 03, CREATED)  
 DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
 DE K+ CHANNEL BETA4 SUBUNIT.  
 OS MUS MUSCULUS (MOUSE).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=BRIN;  
 RC MEDLINE; 96421640.  
 RA FINK M., DUPRAT F., LESAGE F., HEURTEAUX C., ROMÉY G., BARHANIN J.,  
 RA LAZDUNSKI M.;  
 RT "A new K+ channel beta subunit to specifically enhance Kv2.2 (CDRK)  
 RT expression.";  
 RL J. BIOL. CHEM. 271:26341-26348(1996).  
 DR EMBL; U65593; G1695272;  
 KW IONIC CHANNEL.  
 SQ SEQUENCE 249 AA; 27749 MW; 8805DBE7 CRC32;  
  
 Query Match 36.3%; Score 73; DB 11; Length 249;  
 Best Local Similarity 76.9%; Pred. No. 8.98e-01;  
 Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
 Db 21 PHPPGLOGSLDRL 33  
 QY 2 PGPPGLOGRLQL 14  
  
 RESULT 8  
 ID O48721 PRELIMINARY; PRT; 145 AA.  
 AC O48721;  
 DT 01-JUN-1998 (TREMREL. 06, CREATED)  
 DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)  
 DE T9J22.21 PROTEIN.  
 GN T9J22.21.  
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
 OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSEY M.L., BRANDON R.C.,  
 RA SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,  
 RA SOMERVILLE C.R., VENTER J.C.;  
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; AC002505; G2739379;  
 SQ SEQUENCE 145 AA; 15355 MW; 660DDF14 CRC32;  
  
 Query Match 35.3%; Score 71; DB 10; Length 145;  
 Best Local Similarity 39.1%; Pred. No. 1.85e+00;  
 Matches 9; Conservative 8; Mismatches 6; Indels 0; Gaps 0;  
  
 Db 116 PEKPGLEGWVESIMEALGAHADS 138  
 QY 2 PGPPGLOGRLQLQANGHAAIG 24  
  
 RESULT 9  
 ID O80740 PRELIMINARY; PRT; 511 AA.  
 AC O80740;  
 DT 01-NOV-1998 (TREMREL. 08, CREATED)  
 DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
 DE T13D8.6 PROTEIN.  
 GN T13D8.6.  
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
 OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA VISOTSKAIA V.S., SCHWARTZ J.R., KWAN A., TORIUMI M., YU G., OJI, O.,  
 RA LIU S., LI J., ARAUJO R., AU M., BRENDEN V., BUEHLER E., CONWAY A.B.,  
 RA CONWAY A.R., DEWAR K., FENG J., KIM C., KURTZ D., LI Y., PALM C.J.,

Matches 8; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 280 LQEVLDSDSNHVPVSLT 296  
||:||||:||||:||||  
QY 11 LQRLQANGNHAAGILT 27

RESULT 14

ID O64673 PRELIMINARY; PRT: 1015 AA.  
AC O64673;  
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
DE F22013.25  
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA SHINN P., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y., SUN H.,  
RA CONWAY A., CONWAY A., KURTZ D., OJI O., SHEN Y.K., TORIUMI M.,  
RA VYSOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,  
RA ECKER J.R.;  
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AC003981; G3063483; -  
SQ SEQUENCE 1015 AA; 111751 MW; 82FA3C3F CRC32;

Query Match 33.8%; Score 68; DB 10; Length 1015;  
Best Local Similarity 50.0%; Pred. No. 5.34e+00;  
Matches 13; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Db 499 PTRPGLRSLKRRKNGNGPTAATILT 524  
| |||:| | || ||||  
QY 2 PGPPGLQRLQRLQANGNHAAGILT 27

RESULT 15

ID Q14980 PRELIMINARY; PRT: 2115 AA.  
AC Q14980;  
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE NUMA PROTEIN.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
OC CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92176231.  
RA YANG C.H., LAMBIE E.J., SNYDER M.;  
RT "NUMA: an unusually long coiled-coil related protein in the mammalian  
nucleus";  
RL J. CELL BIOL. 116:1303-1317(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94013066.  
RA MAEKAWA T., KURIYAMA R.;  
RT "Primary structure and microtubule-interacting domain of the SP-H  
RT antigen: a mitotic MAP located at the spindle pole and characterized  
RT as a homologous protein to Numa";  
RL J. CELL SCI. 105:589-600(1993).  
DR EMBL; Z11583; G35119; -  
FT CONFLICT 124 124 Q -> P (IN REF. 2).  
FT CONFLICT 1587 1587 Q -> H (IN REF. 2).  
SQ SEQUENCE 2115 AA; 238273 MW; 81A36BA3 CRC32;

Query Match 33.8%; Score 68; DB 4; Length 2115;  
Best Local Similarity 50.0%; Pred. No. 5.34e+00;  
Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 761 RAGRKGLRQLQGLGEAH 778  
|:| ||:||||:|:

QY 1 RPPGPGQLQRLQRLQAN 18

Search completed: Fri Aug 20 21:23:53 1999  
Job time : 41 secs.

\*\*\*\*\*

WORLD

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:14:33 1999; MasPar time 8.03 Seconds  
325.686 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-938-548B-10  
Description: (1-123) from US08938548B.pep  
Perfect Score: 899  
Sequence: 1 VPWAAVTLTLLLLPPALLS.....GRGCPVTVTALAPRGSGV 123

Scoring table: PAM 150  
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-geneseq35  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
14:part14 15:part15 16:part16 17:part17 18:part18  
19:part19 20:part20 21:part21 22:part22 23:part23  
24:part24 25:part25 26:part26 27:part27 28:part28  
29:part29 30:part30 31:part31 32:part32 33:part33  
34:part34 35:part35 36:part36 37:part37 38:part38  
39:part39

Statistics: Mean 29.551; Variance 129.928; scale 0.227

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	899	100.0	123	33	Mouse HFGAN72 recepto	1.08e-73
2	894	99.4	130	30	Mouse hypocretin 35.	3.25e-73
3	853	94.9	130	30	Rat hypocretin 35.	2.79e-69
4	853	94.9	130	33	Rat HFGAN72 receptor	2.79e-69
5	734	81.6	131	33	Human HFGAN72 recepto	6.74e-58
6	113	12.6	378	39	Glial cell line-deriv	4.05e-01
7	113	12.6	400	39	Glial cell line-deriv	4.05e-01
8	113	12.6	400	39	ADNFR-alpha-related	4.05e-01
9	113	12.6	400	32	Human GDNF alpha-3 re	4.05e-01
10	113	12.6	400	32	Human GDNF alpha-3 re	4.05e-01
11	113	12.6	400	28	Human Ret ligand RetL	4.05e-01
12	96	10.7	556	28	Human fibulin type 1	7.98e-00
13	96	10.7	566	2	Fibulin A.	7.98e-00
14	96	10.7	601	2	Fibulin B.	7.98e-00
15	96	10.7	601	28	Human fibulin type 1	7.98e-00
16	96	10.7	683	2	Fibulin C.	7.98e-00

17	96	10.7	683	28	W27600	Human fibulin type 1	7.98e+00
18	96	10.7	703	28	W27601	Human fibulin type 1	7.98e+00
19	91	10.1	440	2	P70134	Natural recombinant h	1.86e+01
20	91	10.1	440	25	W24789	Human lecithin-choles	1.86e+01
21	90	10.0	566	30	W47029	Human N-proteinase (7	2.20e+01
22	90	10.0	954	33	W48978	Mature human chordin	2.20e+01
23	90	10.0	1211	30	W47028	Human N-proteinase (1	2.20e+01
24	89	9.9	58	37	W75074	Human secreted protei	2.60e+01
25	88	9.8	226	37	W72902	Mycobacterium tubercu	3.08e+01
26	88	9.8	254	12	R64190	Human 4-1BB-L polypep	3.08e+01
27	88	9.8	254	25	W26657	Human 4-1BB ligand.	3.08e+01
28	87	9.7	111	7	R35515	Tryptophan aporepress	3.63e+01
29	87	9.7	438	31	W57042	Human aspartic protea	3.63e+01
30	87	9.7	451	34	W54877	Human napsin A protei	3.63e+01
31	87	9.7	599	4	R21690	Prostaglandin endopep	3.63e+01
32	85	9.5	81	37	W57082	Human secreted protei	5.05e+01
33	85	9.5	363	23	W12414	Porcine complement in	5.05e+01
34	84	9.3	35	1	P94256	Truncated E protein f	5.95e+01
35	84	9.3	91	31	W58899	Plasmid pCSJ bacterio	5.95e+01
36	84	9.3	91	31	W58896	Plasmid pAW12/10 PhiX	5.95e+01
37	84	9.3	234	16	R82605	Eph transmembrane tyr	5.95e+01
38	84	9.3	238	13	R71481	Human hek-L protein.	5.95e+01
39	84	9.3	331	33	W48748	Human granulocytic eh	5.95e+01
40	84	9.3	1001	30	W43023	Mouse protein tyrosin	5.95e+01
41	84	9.3	1239	9	R45945	Glutamic acid recepto	5.95e+01
42	83	9.2	179	8	R50056	ICP34.5 fragment.	7.01e+01
43	83	9.2	235	39	W67769	Human flt3-ligand.	7.01e+01
44	83	9.2	235	13	R66175	Human S86/S109 Flt3 l	7.01e+01
45	83	9.2	551	13	R77858	S. clavuligerus ORF1	7.01e+01

ALIGNMENTS

RESULT 1  
ID W61383 standard; Protein; 123 AA.  
AC W61383;  
DT 02-OCT-1998 (first entry)  
DE Mouse HFGAN72 receptor protein.  
KW HFGAN72 receptor; eating disorders; renal disease; heart failure;  
KW sexual dysfunction; HIV; cancer; Parkinson's disease; urinary retention;  
KW osteoporosis; angina pectoris; myocardial infarction; psychotic disorder;  
KW neurological disorder.  
OS Mus sp.  
FH Key  
FT Region 33..65 Location/Qualifiers  
FT /note= "Ligand 72A"  
FT Region 69..96  
FT /note= "Ligand 72B"  
FN EP-849361-A2.  
PD 24-JUN-1998.  
PF 17-DEC-1997; 310216.  
PR 26-SEP-1997; US-939093.  
PR 17-DEC-1996; US-033604.  
PR 19-MAR-1997; US-820519.  
PR 02-JUL-1997; US-887382.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PI Bergsma DJ Brooks DP, Gellai M, Wilson S, Yanagisawa M;  
PI WPI; 98-324672/29.  
DR HFGAN72 receptor ligands - and corresponding DNA, agonists,  
PT antibodies, antagonists, etc.  
PS Claim 5: Fig 5; 35pp; English.  
CC The HFGAN72 receptor protein contains two ligands whose antagonists can be  
CC used for treating obesity, diabetes, anorexia nervosa, bulimia, cachexia,  
CC chronic renal failure, renal disease, congestive heart failure, impaired  
CC glucose tolerance and sexual dysfunction. The agonist is useful for  
CC treating anorexia nervosa, bulimia and cachexia. The HFGAN72 receptor  
CC ligand is useful for treating e.g. bacterial, fungal, protozoan and viral  
CC infections, particularly infections caused by HIV-1 or HIV-2, pain,  
CC cancers, anorexia nervosa, bulimia, cachexia, obesity, diabetes, asthma,  
CC Parkinson's disease, both acute and congestive heart failure,  
CC hypotension, hypertension, urinary retention, osteoporosis, angina  
CC pectoris, myocardial infarction, ulcers, asthma, allergies, benign



```

PD 24-JUN-1998.
PF 17-DEC-1997; 310216.
PR 26-SEP-1997; US-939093.
PR 17-DEC-1996; US-033604.
PR 19-MAR-1997; US-820519.
PR 02-JUL-1997; US-887382.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
PI Bergsma DJ, Brooks DP, Gellai M, Wilson S, Yanagisawa M;
DR WPI: 98-324672/29.
DR N-PSDB; V28139.
PT HFGAN72 receptor ligands - and corresponding DNA, agonists,
PT antibodies, antagonists, etc.
PS Claim 5; Fig 4: 35pp; English.
CC The HFGAN72 receptor protein contains two ligands whose antagonists can
CC be used for treating obesity, diabetes, anorexia nervosa, bulimia, cachexia,
CC chronic renal failure, renal disease, congestive heart failure,
CC glucose tolerance and sexual dysfunction. The agonist is useful for
CC treating anorexia nervosa, bulimia and cachexia. The HFGAN72 receptor
CC ligand is useful for treating e.g. bacterial, fungal, protozoan and viral
CC infections, particularly infections caused by HIV-1 or HIV-2, pain,
CC cancers, anorexia nervosa, bulimia, cachexia, obesity, diabetes,
CC Parkinson's disease, both acute and congestive heart failure,
CC hypotension, hypertension, urinary retention, osteoporosis, angina
CC pectoris, myocardial infarction, ulcers, asthma, allergies, benign
CC prostatic hypertrophy, chronic renal failure, renal disease, impaired
CC glucose tolerance, sexual dysfunction and psychotic and neurological
CC disorders including anxiety, schizophrenia, manic depression, delirium,
CC dementia, severe mental retardation and dyskinesias such as Huntington's
CC disease or Gilles de la Tourette's syndrome.
SQ Sequence 130 AA;

Query Match 94.9%; Score 853; DB 33; Length 130;
Best Local Similarity 95.1%; Pred. No. 2.79e-69;
Matches 117; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 8 vpxaavtllllllppallsigvdaqlpdcrcqktcsrlyellhgnhaagiltlgk 67
QY 1 VPAAVATLLLLLLPPALLSLGVDAQPLPDCRCQKTCSRLYELLHGNHGAAGILTLG 60

Db 68 rrpqpgglqgrlrlqgngnhaagiltmrragaelpypcpgrcrtatatlaprgg 127
QY 61 RRPQPGGLQGRRLRLQNGNHAAGILTMRRAGAELEPHPCSGRGCTVTITTALAPRG 120

Db 128 srj 130
QY 121 SGV 123

RESULT 5
ID W61381 standard; Protein; 131 AA.
AC W61381;
DT 02-OCT-1998 (first entry)
DE Human HFGAN72 receptor protein.
KW HFGAN72 receptor; eating disorders; renal disease; heart failure;
KW sexual dysfunction; HIV; cancer; Parkinson's disease; urinary retention;
KW osteoporosis; angina pectoris; myocardial infarction; psychotic disorder;
KW neurological disorder.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Region 34..66
FT /note= "Ligand 72A"
FT Region 70..97
FT /note= "Ligand 72B"
FT
FT
PN EP-849361-A2.
PD 24-JUN-1998.
PF 17-DEC-1997; 310216.
PR 26-SEP-1997; US-939093.
PR 17-DEC-1996; US-033604.
PR 19-MAR-1997; US-820519.
PR 02-JUL-1997; US-887382.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.

PD 24-JUN-1998.
PF 17-DEC-1997; 310216.
PR 26-SEP-1997; US-939093.
PR 17-DEC-1996; US-033604.
PR 19-MAR-1997; US-820519.
PR 02-JUL-1997; US-887382.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.

PI Bergsma DJ, Brooks DP, Gellai M, Wilson S, Yanagisawa M;
DR WPI: 98-324672/29.
DR N-PSDB; V28138.
PT HFGAN72 receptor ligands - and corresponding DNA, agonists,
PT antibodies, antagonists, etc.
PS Claim 2; Fig 2: 35pp; English.
CC The HFGAN72 receptor protein contains two ligands whose antagonists can
CC be used for treating obesity, diabetes, anorexia nervosa, bulimia,
CC cachexia, chronic renal failure, renal disease, congestive heart failure,
CC impaired glucose tolerance and sexual dysfunction. The agonist is
CC useful for treating anorexia nervosa, bulimia and cachexia. The HFGAN72
CC receptor ligand is useful for treating e.g. bacterial, fungal, protozoan
CC and viral infections, particularly infections caused by HIV-1 or HIV-2,
CC pain, cancers, anorexia nervosa, bulimia, cachexia, obesity, diabetes,
CC asthma, Parkinson's disease, both acute and congestive heart failure,
CC hypotension, hypertension, urinary retention, osteoporosis, angina
CC pectoris, myocardial infarction, ulcers, asthma, allergies, benign
CC prostatic hypertrophy, chronic renal failure, renal disease, impaired
CC glucose tolerance, sexual dysfunction and psychotic and neurological
CC disorders including anxiety, schizophrenia, manic depression, delirium,
CC dementia, severe mental retardation and dyskinesias such as Huntington's
CC disease or Gilles de la Tourette's syndrome.
SQ Sequence 131 AA;

Query Match 81.6%; Score 734; DB 33; Length 131;
Best Local Similarity 81.5%; Pred. No. 6.74e-58;
Matches 101; Conservative 12; Mismatches 10; Indels 1; Gaps 1;

Db 8 vswaavtllllllppallsigvdaqlpdcrcqktcsrlyellhgnhaagiltlg 67
QY 1 VPAAVATLLLLLLPPALLSLGVDAQPLPDCRCQKTCSRLYELLHGNHGAAGILTLG 59

Db 68 krrsgpgglqgrlrlqasgnhaagiltmrragaelpcprcgrcsapaasvapg 127
QY 60 KRRSGPGGLQGRRLRLQANGNHAAGILTMRRAGAELEPHPCSGRGCTVTITTALAPRG 119

Db 128 qsggi 131
QY 120 GSGV 123

RESULT 6
ID W84185 standard; Protein; 378 AA.
AC W84185;
DT 25-MAR-1999 (first entry)
DE Glial cell line-derived neurotrophic factor receptor gamma 1.
KW Glial cell line-derived neurotrophic factor receptor gamma 1;
KW GDNFR-alpha; glial cell line-derived neurotrophic factor; GDNF;
KW neurodegenerative disease; amyotrophic lateral sclerosis; GDNFR-gamma1;
KW Parkinson's disease; schizophrenia; insomnia; tardive dyskinesia;
KW hypertension; pituitary adenomas; hyperprolactinemia; thyroid tumour;
KW renal disorder; kidney failure; gut dysfunction; regeneration;
KW cardiomyocyte; epithelium; hepatocyte.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..31
FT /label= signal_peptide
FT Protein 32..378
FT /label= mature_protein
FT Domain 32..360
FT /note= "extracellular domain"
FT Domain 361..378
FT /note= "transmembrane domain"
FT
FT
PN W09853069-A2.
PD 26-NOV-1998.
PF 20-MAY-1998; U10328.
PR 27-JUN-1997; US-884638.
PR 20-MAY-1997; US-047092.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Gentz RL, Hsu T, Ni J, Ruben SM, Young P;
DR WPI: 99-070150/06.
DR N-PSDB; V99333.
PT New isolated glial cell derived neurotrophic factor receptors - used

```

CC This sequence represents a novel glial cell line-derived neurotrophic factor alpha-3 receptor (GDNF alpha-3). This protein can be used to treat e.g. neurodegenerative diseases (such as Parkinson's Disease, CC anyotrophic lateral sclerosis (ALS), spinal muscular atrophy (SMA), CC Huntington's Disease, Alzheimer's Disease, diabetic neuropathy), CC muscular diseases (including the muscular dystrophies) and nerve and CC muscle trauma and in diagnostic assays for such conditions.

SQ Sequence 400 AA;

Query Match 12.6%; Score 113; DB 32; Length 400;  
Best Local Similarity 47.4%; Pred. NO. 4.05e-01;  
Matches 18; Conservative 9; Mismatches 9; Indels 2; Gaps

Db 9 plppvmllllppsplpaa-gdplptesrlmn-sc 44  
|:::|:|||||||::|:::|::|:|:  
Qy 2 PWAATLELLLPALLSLGVDAQPLPDCCRQRTCS 39

RESULT 11

ID W37463 standard; Protein; 400 AA.  
AC W37463;  
DE 21-MAY-1998 (first entry)  
DT Human Ret ligand RetL3.  
KW Ret ligand; RetL; RetL3; receptor; signal transduction; human;  
KW cell growth; renal cell; nerve cell; renal failure; nephritis;  
KW kidney transplant; toxic injury; hypoxic injury;  
KW neurodegeneration; motor neurone disease; multiple sclerosis;  
KW infection; meningitis; myelopathy; Creutzfeldt-Jakob disease;  
KW cranial nerve injury; spinal cord injury; Down's syndrome;  
KW cerebral palsy; Lyme disease; muscular dystrophy;  
KW myasthenia gravis; tumour; therapy.

OS Homo sapiens.  
PN WO9744356-A2.  
PD 27-NOV-1997.  
PF 07-MAY-1997. H07726

10-APR-1997; US-017427.  
PR  
PR 08-MAY-1996; US-017427.  
PR 07-JUN-1996; US-019300.  
PR 16-JUL-1996; US-021859.  
PR (BIOJ ) BIOGEN INC.  
PA Cate RL, Hession C, Sanicola-Nadel M;  
PI WPI: 98-018431/02.  
DR N-PSDB: V00251  
DR

PS Claim 2: Page 85-86; 113pp: English.

CC This amino acid sequence comprises human Ret ligand (RetL) RetL3,

CC deduced from cDNA clones (see V00251) isolated from a adult heart

CC and spinal cord libraries. Rat and human RetL1, human RetL2 and

CC mouse RetL3 sequences (see W37457-62) are also claimed. Human

CC RetL3 is 34.3% identical to human RetL1, 34.9% identical to human

CC RetL2 and 76.8% identical to murine RetL3. Ret ligand is a key

CC component of the Ret signalling pathway that specifically

CC interacts with Ret receptor protein, triggering Ret dimerisation

CC and/or autophosphorylation of the Ret tyrosine kinase domain.

CC Vectors containing retL3 DNA and prokaryotic or eukaryotic host

CC cells transformed or transfected with these vectors are claimed, as

CC well as a method for production of RetL3, its soluble variants and

CC fusion proteins with a toxin, imageable compound or radionuclide.

CC RetL3, optionally when expressed from vectors in vivo, is used to

CC promote growth of new tissue and survival of damaged tissue,

CC particularly kidney or neural tissue. Typical applications are in

CC renal failure, nephritis, kidney transplants, toxic or hypoxic

CC injury, neurodegeneration, motor neurone disease, multiple sclerosis

CC bacterial, viral or prion infections (e.g. meningitis, myelopathy

CC associated with HIV or Creutzfeldt-Jakob disease), cranial nerve or

CC spinal cord injury, developmental disorders such as Down's syndrome

CC and cerebral palsy, or conditions involving the peripheral nervous

CC system (lyme disease, muscular dystrophy and myasthenia gravis).

CC Fusion proteins are used to deliver toxins etc. to Ret-expressing

CC cells, especially tumours.

CC single residue. Embodied within repeats 5, 6, 7 and 8 is the  
 CC consensus sequence for Asp and Asn hydroxylation. The 7th repeat  
 CC contains the consensus O-glycosylation sequence. CXCXPC.  
 CC Immediately following each repeat is a pentapeptide with the  
 CC consensus sequence XD(I/V)(D/N)E. Fibulin binds to the cyto-  
 CC plasmic domain of the betal subunit of integrin adhesion  
 CC receptors in a cation-dependent, EDTA-reversible manner. It can  
 CC be used to manipulate adhesion of cells to fibronectin, collagen,  
 CC laminin, and possibly also other proteins. Antibodies reactive  
 CC with the protein have important diagnostic and therapeutic uses.  
 CC See also R1147, R1149 and R1150.  
 SQ Sequence 566 AA;

Query Match 10.7%; Score 96; DB 2: Length 566;

Best Local Similarity 57.1%; Pred. No. 7.98e+00;

Matches 16; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

Db 10 vplp1lllglllaagvdadvleacc 37

QY 6 VTLLLLLLPPALLSLGVDQAQPLPD-CC 32

RESULT 14

ID R1149 standard; Protein; 601 AA.

AC R1149;

DT 21-MAY-1991 (first entry)

DE Fibulin B.

KW Beta-1 integrin; adhesion; receptor; fibronectin.

OS Homo sapiens.

FH Key Location/Qualifiers

FT peptide 1..29

FT protein /label= signal sequence

FT modified\_site 30..601

FT /label= fibulin B

FT modified\_site 98

FT /label= N-linked glycosylation

FT modified\_site 535

FT modified\_site 339

FT modified\_site 36.144

FT region /label= type I motif

FT region 36..59

FT region /label= repeat unit 1

FT region 112..144

FT region /label= repeat unit 2

FT region 144..179

FT region /label= Glu/Asp-rich region

FT region 180..566

FT region /label= type II motif

FT region 180..214

FT region /label= repeat unit 1

FT region 215..219

FT region /label= consensus pentapeptide

FT region 220..260

FT region /label= repeat unit 2

FT region 261..265

FT region /label= consensus pentapeptide

FT region 266..306

FT region /label= repeat unit 3

FT region 307..311

FT region /label= consensus pentapeptide

FT region 312..354

FT region /label= repeat unit 4

FT region 355..359

FT region /label= consensus pentapeptide

FT region 360..397

FT region /label= repeat unit 5

FT region 398..402

FT region /label= consensus pentapeptide

FT region 403..439

FT region /label= repeat unit 6

FT region 440..444

FT region

FT region

FT region

FT region

FT region

FT region

FT region /label= consensus pentapeptide

FT 445..479

FT /label= repeat unit 7

FT 480..484

FT /label= consensus pentapeptide

FT 485..529

FT /label= repeat unit 8

FT 530..535

FT /label= consensus pentapeptide

PN W09102755-A.

PD 07-MAR-1991.

PF 17-AUG-1990; U04662.

PR 16-AUG-1989; US-395773.

PA (LJOL-) LA JOLLA CANCER RES.

PA (AMNA-) AMER NAT RED CROSS.

PI Ruoslahti EI, Argraves WS;

DR WPI; 91-087250/12.

DR N-PSDB; Q11009.

PT Purified fibulin, DNA encoding it and antibodies reactive with it

PT - useful as diagnostic and therapeutic component.

PS Claim 10; Fig 4; 56pp; English.

CC The fibulin A, B and C forms are identical from their N-terminals

CC to a divergence point at posn. 566 (terminal codon in fibulin A)

CC after which they are distinct, encoding polypeptides of 566, 601

CC and 683 residues resp. All three forms are rich in cysteine (11 %)

CC and analysis wrt no. and spacing of the Cys residues revealed two

CC types of repeat motif (I and II). The type I motif, CC(X)12C-

CC (X)9-10C(X)6CC is repeated twice, separated by an imperfect form of

CC the repeat lacking two Cys residues. The same motif is found in

CC complement components C3a, C4a and C5a; the inverse is found in

CC albumin, vitamin D-binding protein and alpha-fetoprotein. The

CC disulphide-stabilized loop structure is thought to be conserved.

CC The type II motif, related to repeats found in epidermal growth

CC factor precursor is a 6-Cys pattern repeated nine times, although

CC the ninth repeat in the A form is incomplete. Four of the repeats,

CC (2, 3, 4 and 9) differ from the EGF-like motif in having a 4-6

CC residue insertion between cysteines 4 and 5 instead of the usual

CC single residue. Embodied within repeats 5, 6, 7 and 8 is the

CC consensus sequence for Asp and Asn hydroxylation. The 7th repeat

CC contains the consensus O-glycosylation sequence, CXCXPC.

CC Immediately following each repeat is a pentapeptide with the

CC consensus sequence XD(I/V)(D/N)E. Fibulin binds to the cyto-

CC plasmic domain of the betal subunit of integrin adhesion

CC receptors in a cation-dependent, EDTA-reversible manner. It can

CC be used to manipulate adhesion of cells to fibronectin, collagen,

CC laminin, and possibly also other proteins. Antibodies reactive

CC with the protein have important diagnostic and therapeutic uses.

CC See also R1147, R1148 and R1150.

CC Sequence 601 AA;

SQ

Query Match 10.7%; Score 96; DB 2: Length 601;

Best Local Similarity 57.1%; Pred. No. 7.98e+00;

Matches 16; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

Db 10 vplp1lllglllaagvdadvleacc 37

QY 6 VTLLLLLLPPALLSLGVDQAQPLPD-CC 32

RESULT 15

ID W27599 standard; Protein; 601 AA.

AC W27599;

DT 12-MAY-1998 (first entry)

DE Human fibulin type I isoform (variant B).

KW Human fibulin protein; met-OB protein; hypertension; obesity;

KW Isoform; type II diabetes; fibrogenesis.

OS Homo sapiens.

PN W09738014-A1.

PD 16-OCT-1997.

PF 03-APR-1997; U06280.

PF 04-APR-1996; US-627636.

PA (AMGE-) AMGEN INC.

PI Bennett LG;

	Query Match	11.8%;	Score 106;	DB 2;	Length 679;
	Best Local Similarity	32.3%;	Pred. NO. 1.65e-02;		
	Matches	20;	Mismatches 22;	Indels 4;	Gaps 4
Db	598 KYCRSK-YETIQGHNDNAADVLEAIKREMPAELL-R-ASLRHTEDEORNFLLNVRSA	654			
	:   :   :   :				
QV	35 KTCSCRLVELLHGAG-NIIAAAGTILTGKKRRPGPPGLOGLRLOLANGNHAAAGILTMGRRA	93			
	:   :   :   :				

```

DATE          19-Apr-1991 #sequence_revision 19-Apr-1991 #text_change
10-Sep-1997
ACCESSIONS    C36346; A36346; A32826
REFERENCE     Argraves, W.S.; Tran, H.; Burgess, W.H.; Dickerson, K.
#authors      J. Cell Biol. (1990) 111:3155-3164
#journal      Fibulin is an extracellular matrix and plasma glycoprotein
#title        with repeated domain structure.
#cross-references MUID:91100426
#accession    C36346
#molecule_type mRNA
#residues     1-683 #label ARG
#cross-references GB:X53743; NID:g31418; PID:g31419
#accession    A36346
#molecule_type mRNA
#residues     1-566 #label AR2
#cross-references GB:X53741; NID:g31414; PID:g31415
REFERENCE     Argraves, W.S.; Dickerson, K.; Burgess, W.H.; Ruoslahti, E.
#authors      Cell (1989) 58:623-629
#journal      Fibulin, a novel protein that interacts with the fibronectin
#title        receptor beta-subunit cytoplasmic domain.
#cross-references MUID:89354537
#accession    A32826
#molecule_type protein
#residues     30-35, 'SX', 38-40, 'SH', 43-44 #label AR3
GENETICS
#gene         GDB:FBLN1; FBLN
#map_position 22q13.3-22q13.3
#keywords     alternative splicing; glycoprotein
FEATURE
1-29          #domain signal sequence #status predicted #label SIG\
30-683        #product fibulin 1 splice form C #status predicted
#label MATN
#binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY
length 683 #molecular-weight 74475 #checksum 7443
Query Match   10.7%; Score 96; DB 2; Length 683;
Best Local Similarity 57.1%; Pred. No. 3.46e-01;
Matches 16; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

Db 10 VPLPLLLLGALLAAGVDADVLEACC 37
| | | | | | | | | | | | | | | | | | | |
QY 6 VTLLLLLLPPALLSLGVDQAQPLPD-CC 32

RESULT 6
ENTRY   #type complete
TITLE   eyeless, long form - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE    09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change
17-Oct-1997
ACCESSIONS I45557
REFERENCE   Quiring, R.; Walldorf, U.; Kloter, U.; Gehring, W.J.
#authors    Science (1994) 265:785-789
#journal     Homology of the eyeless gene of Drosophila to the Small eye
#title       gene in mice and Aniridia in humans [see comments].
#cross-references MUID:94323757
#accession  I45557
#status     preliminary
#molecule_type mRNA
#residues   1-838 #label RES
#cross-references EMBL:X79493; NID:g641809; PID:g641810
GENETICS
#gene       FlyBase:ey
#cross-references FlyBase:FBgn0005558
#introns    37/1; 92/2; 152/3; 371/1; 429/1; 521/1; 639/2
CLASSIFICATION #superfamily unassigned homeobox proteins; homeobox homology;
paired box homology
KEYWORDS      alternative splicing; DNA binding; homeobox; nucleus;

19-Apr-1991 #sequence_revision 19-Apr-1991 #text_change
10-Sep-1997
ACCESSIONS    C36346; A36346; A32826
REFERENCE     Argraves, W.S.; Tran, H.; Burgess, W.H.; Dickerson, K.
#authors      J. Cell Biol. (1990) 111:3155-3164
#journal      Fibulin is an extracellular matrix and plasma glycoprotein
#title        with repeated domain structure.
#cross-references MUID:91100426
#accession    C36346
#molecule_type mRNA
#residues     1-683 #label ARG
#cross-references GB:X53743; NID:g31418; PID:g31419
#accession    A36346
#molecule_type mRNA
#residues     1-566 #label AR2
#cross-references GB:X53741; NID:g31414; PID:g31415
REFERENCE     Argraves, W.S.; Dickerson, K.; Burgess, W.H.; Ruoslahti, E.
#authors      Cell (1989) 58:623-629
#journal      Fibulin, a novel protein that interacts with the fibronectin
#title        receptor beta-subunit cytoplasmic domain.
#cross-references MUID:89354537
#accession    A32826
#molecule_type protein
#residues     30-35, 'SX', 38-40, 'SH', 43-44 #label AR3
GENETICS
#gene         GDB:FBLN1; FBLN
#map_position 22q13.3-22q13.3
#keywords     alternative splicing; glycoprotein
FEATURE
1-29          #domain signal sequence #status predicted #label SIG\
30-683        #product fibulin 1 splice form C #status predicted
#label MATN
#binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY
length 683 #molecular-weight 74475 #checksum 7443
Query Match   10.7%; Score 96; DB 2; Length 683;
Best Local Similarity 57.1%; Pred. No. 3.46e-01;
Matches 16; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

Db 10 VPLPLLLLGALLAAGVDADVLEACC 37
| | | | | | | | | | | | | | | | | | | |
QY 6 VTLLLLLLPPALLSLGVDQAQPLPD-CC 32

RESULT 6
ENTRY   #type complete
TITLE   eyeless, long form - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE    09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change
17-Oct-1997
ACCESSIONS I45557
REFERENCE   Quiring, R.; Walldorf, U.; Kloter, U.; Gehring, W.J.
#authors    Science (1994) 265:785-789
#journal     Homology of the eyeless gene of Drosophila to the Small eye
#title       gene in mice and Aniridia in humans [see comments].
#cross-references MUID:94323757
#accession  I45557
#status     preliminary
#molecule_type mRNA
#residues   1-838 #label RES
#cross-references EMBL:X79493; NID:g641809; PID:g641810
GENETICS
#gene       FlyBase:ey
#cross-references FlyBase:FBgn0005558
#introns    37/1; 92/2; 152/3; 371/1; 429/1; 521/1; 639/2
CLASSIFICATION #superfamily unassigned homeobox proteins; homeobox homology;
paired box homology
KEYWORDS      alternative splicing; DNA binding; homeobox; nucleus;

transcription regulation
#domain paired box homology #label PBH\
#domain homeobox homology #label HOX
#length 838 #molecular-weight 87662 #checksum 392
SUMMARY
Query Match   10.2%; Score 92; DB 2; Length 838;
Best Local Similarity 28.3%; Pred. No. 1.11e-00;
Matches 15; Conservative 19; Mismatches 17; Indels 2; Gaps 2;

Db 681 PPMAPAHHIVPGDGRPAGV-GLGSGQSANLGA-SCSGSYEVLSAYALPP 731
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 PPGLOQLRQLQANQNHAGILTMRRAGAELEPHPCSGRCPTVTITLALP 117

RESULT 7
ENTRY   #type complete
TITLE   phosphatidylcholine--sterol O-acyltransferase (EC 2.3.1.43)
ALTERNATE_NAMES lecitithin--cholesterol acyltransferase precursor;
phospholipid--cholesterol acyltransferase precursor
ORGANISM #formal_name Homo sapiens #common_name man
DATE      04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change
02-Sep-1997
ACCESSIONS A00571; A25575; A29661; J00036; A29133; I52260; A28511
REFERENCE   McLean, J.; Fielding, C.; Drayna, D.; Dieplinger, H.; Baer,
#authors      B.; Kohr, W.; Henzel, W.; Lawn, R.
#journal      Proc. Natl. Acad. Sci. U.S.A. (1986) 83:2335-2339
#title        Cloning and expression of human lecithin-cholesterol
#cross-references MUID:86205950
#accession    A00571
#molecule_type mRNA
#residues     1-440 #label MCL1
#cross-references GB:M12625; NID:g187022; PID:g307117
REFERENCE     McLean, J.; Wion, K.; Drayna, D.; Fielding, C.; Lawn, R.
#authors      Nucleic Acids Res. (1986) 14:9397-9406
#journal      Human lecithin-cholesterol acyltransferase gene: complete
#title        gene sequence and sites of expression.
#cross-references MUID:87091568
#accession    A25575
#molecule_type DNA
#residues     1-440 #label MCL2
#cross-references GB:X04981; NID:g34286; PID:g34287
REFERENCE     Rogne, S.; Skretting, G.; Larsen, F.; Myklebost, O.; Mevag,
#authors      B.; Carlson, L.A.; Holmquist, L.; Gjone, E.; Prydz, H.
#journal      Biochem. Biophys. Res. Commun. (1987) 148:161-169
#title        The isolation and characterisation of a cDNA clone for human
#cross-references MUID:88050946
#accession    J00036
#molecule_type mRNA
#residues     17-256, 'H', 258-440 #label TAT
#cross-references GB:M17959; NID:g187026; PID:g386858
REFERENCE     Tata, F.; Chaves, M.E.; Markham, A.F.; Scraze, G.D.;
#authors      Waterfield, M.D.; McIntyre, N.; Williamson, R.; Humphries,
#journal      S.E.
#title        Biochim. Biophys. Acta (1987) 910:142-148
#cross-references MUID:88050946
#accession    J00036
#molecule_type mRNA
#residues     17-256, 'H', 258-440 #label TAT
#cross-references GB:X06537; NID:g34284; GB:M26268; NID:g187024;
#note         the authors translated the codon CAT for residue 241 as

```

TITLE	natriuretic peptide receptor A precursor - human
ALTERNATE_NAMES	atrial natriuretic peptide receptor; atrionatriuretic peptide receptor A; guanylate cyclase A, membrane-bound guanylate cyclase (EC 4.6.1.2) A
CONTAINS	
ORGANISM	#formal_name Homo sapiens #common_name man
DATE	30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 26-Feb-1999
ACCESSIONS	S04459; I52846
REFERENCE	S04459
#authors	Lowe, D.G.; Chang, M.S.; Hellmiss, R.; Chen, E.; Singh, S.; Garbers, D.L.; Goeddel, D.V.
#journal	EMBO J. (1989) 8:1377-1384
#title	Human atrial natriuretic peptide receptor defines a new paradigm for second messenger signal transduction.
#cross-references	MUID:89356605
#accession	S04459
#molecule_type	mRNA
#residues	1-1061 #label LOW
#cross-references	EMBL:X15357; NID:g28229; PTD:g28230 I52846
REFERENCE	
#authors	Pardhasaradhi, K.; Kuty, R.K.; Gentleman, S.; Krishna, G.
#journal	Cell. Mol. Neurobiol. (1994) 14:1-7
#title	Expression of mRNA for atrial natriuretic peptide receptor guanylate cyclase (ANPRA) in human retina.
#cross-references	MUID:95042574
#accession	I52846
#status	preliminary; translated from GB/EMBL/DDBJ
#molecule_type	mRNA
#residues	634-1048 #label RES
#cross-references	GB:S72628; NID:g619241
#experimental_source	retina
GENETICS	
#gene	GDB:NPRL1; NPRA; ANPRA
#map_position	1q21-1q22
CLASSIFICATION	#superfamily membrane-bound guanylate cyclase; guanylate cyclase catalytic domain homology; natriuretic peptide-binding domain homology; protein kinase homology ATP; carbon-oxygen lyase; cGMP biosynthesis; glycoprotein; hormone receptor; phosphorus-oxygen lyase; transmembrane protein
KEYWORDS	
FEATURE	
1-32	#domain signal sequence #status predicted #label SIG\
33-1061	#product natriuretic peptide receptor A #status predicted #label MAT\
33-473	#domain extracellular #status predicted #label EXT\
113-453	#domain natriuretic peptide-binding domain homology #label NPB\
474-494	#domain transmembrane #status predicted #label TM\
495-1061	#domain intracellular #status predicted #label INT\
526-808	#domain protein kinase homology #label KIN\
828-1055	#domain guanylate cyclase catalytic domain homology #label CAT\
34,45,212,338,379,	#binding_site carbohydrate (Asn) (covalent) #status predicted
386,427	
SUMMARY	#length 1061 #molecular-weight 118918 #checksum 6297
Query Match	9.9%; Score 89; DB 1; Length 1061;
Best Local Similarity	68.2%; Pred. No. 2,61e+00;
Matches	15; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
Db	14 LLLLLLPPLLLLRGSHAGNL 35
Qy	8 LLLLLLPPALLSL-GVDAQPL 28
RESULT	11
ENTRY	T00382 #type fragment
TITLE	hypothetical protein KIAA0634 - human (fragment)
ORGANISM	#formal_name Homo sapiens #common_name man
DATE	01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 01-Feb-1999

\*\*\*\*\*

W P E R E H  
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:18:21 1999; MasPar time 2.85 Seconds

Tabular output not generated. 437.587 Million cell updates/sec

Title: >US-08-938-548B-10  
Description: (1-123) from US08938548B.pep  
Perfect Score: 899  
Sequence: 1 VPWAAVTLTLLLLPPALLS.....GRGCPVTVTALAPRGSGV 123

Scoring table: PAM 150  
Gap 11

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: a-issued  
1:5A\_COMB 2:5B\_COMB 3:PCIT9\_COMB 4:backfiles1

Statistics: Mean 27.574; Variance 124.910; scale 0.221

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	90	10.0	954	2	US-08-749-Sequence 3, Applicatio	9.49e+00
2	88	9.8	254	1	US-08-236-Sequence 4, Applicatio	1.32e+01
3	84	9.3	234	1	US-08-299-Sequence 5, Applicatio	2.57e+01
4	84	9.3	238	1	US-08-453-Sequence 2, Applicatio	2.57e+01
5	84	9.3	238	1	US-08-240-Sequence 3, Applicatio	2.57e+01
6	84	9.3	1239	1	US-08-026-Sequence 6, Applicatio	3.02e+01
7	83	9.2	235	1	US-08-993-Sequence 6, Applicatio	3.02e+01
8	83	9.2	235	1	US-08-243-Sequence 6, Applicatio	3.02e+01
9	83	9.2	235	3	PCIT-US94-0Sequence 2, Applicatio	3.02e+01
10	83	9.2	263	3	PCIT-US91-0Sequence 2, Applicatio	3.02e+01
11	82	9.1	220	4	5175255-4Patent No. 5175255.	3.56e+01
12	82	9.1	3729	2	US-08-804-Sequence 4, Applicatio	3.56e+01
13	81	9.0	226	4	5498600-2Patent No. 5175255.	4.18e+01
14	81	9.0	241	4	5175255-8Patent No. 5175255.	4.18e+01
15	81	9.0	241	4	5175255-2Patent No. 5175255.	4.18e+01
16	81	9.0	282	1	US-08-445-Sequence 1, Applicatio	4.18e+01
17	80	8.9	492	1	US-08-001-Sequence 4, Applicatio	4.92e+01
18	80	8.9	492	1	US-07-794-Sequence 4, Applicatio	4.92e+01
19	80	8.9	684	2	US-08-555-Sequence 12, Applicati	4.92e+01
20	80	8.9	1337	3	PCIT-US95-0Sequence 2, Applicatio	4.92e+01
21	79	8.8	241	1	US-08-387-Sequence 4, Applicatio	5.78e+01
22	79	8.8	241	4	5194596-15Patent No. 5194596.	5.78e+01
23	79	8.8	241	3	PCIT-US96-0Sequence 9, Applicatio	5.78e+01

24	79	8.8	3170	2	US-07-642-Sequence 5, Applicatio	5.78e+01
25	79	8.8	4550	2	US-08-804-Sequence 8, Applicatio	5.78e+01
26	78	8.7	442	2	US-08-893-Sequence 4, Applicatio	6.78e+01
27	78	8.7	442	2	US-08-893-Sequence 2, Applicatio	6.78e+01
28	78	8.7	555	1	US-08-484-Sequence 6, Applicatio	6.78e+01
29	78	8.7	1184	1	US-08-446-Sequence 20, Applicati	6.78e+01
30	78	8.7	1184	2	US-08-805-Sequence 20, Applicati	6.78e+01
31	78	8.7	1184	2	US-08-064-Sequence 20, Applicati	6.78e+01
32	78	8.7	1184	1	US-08-446-Sequence 20, Applicati	6.78e+01
33	78	8.7	1187	1	US-08-097-Sequence 13, Applicati	6.78e+01
34	78	8.7	1187	1	US-08-357-Sequence 8, Applicatio	6.78e+01
35	78	8.7	1187	3	PCIT-US95-1Sequence 8, Applicatio	6.78e+01
36	77	8.6	589	1	US-07-668-Sequence 2, Applicatio	7.96e+01
37	77	8.6	589	1	US-08-429-Sequence 2, Applicatio	7.96e+01
38	77	8.6	589	3	PCIT-US91-0Sequence 2, Applicatio	7.96e+01
39	77	8.6	613	2	US-08-465-Sequence 2, Applicatio	7.96e+01
40	77	8.6	613	3	PCIT-US94-1Sequence 2, Applicatio	7.96e+01
41	77	8.6	1865	2	US-08-588-Sequence 2, Applicatio	7.96e+01
42	77	8.6	1865	2	US-08-971-Sequence 2, Applicatio	7.96e+01
43	77	8.6	2509	1	US-08-469-Sequence 10, Applicati	7.96e+01
44	76	8.5	231	3	PCIT-US95-0Sequence 6, Applicatio	9.33e+01
45	76	8.5	529	3	PCIT-US94-0Sequence 2, Applicatio	9.33e+01

## ALIGNMENTS

RESULT 1  
ID US-08-749-169A-3 STANDARD; PRT; 954 AA.

XX  
AC  
XX  
XX  
DT

Sequence 3, Application US/08749169A

Sequence 3, Application US/08749169A

Patent No. 5846770

GENERAL INFORMATION:

APPLICANT: RACIE, Lisa

APPLICANT: LAVALLIE, Edward

APPLICANT: DEROBERTIS, Edward

TITLE OF INVENTION: CHORDIN COMPOSITIONS

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: USA

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/749,169A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: LAZAR, Steven R.

REGISTRATION NUMBER: 32,618

REFERENCE/DOCKET NUMBER: GI 5284

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8260

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 954 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 954 AA; 101960 MW; 4424673 CN;

SQ

CC TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
CC TITLE OF INVENTION: RECEPTOR HEK  
CC NUMBER OF SEQUENCES: 4  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: IMMUNEX CORPORATION  
CC STREET: 51 UNIVERSITY STREET  
CC CITY: SEATTLE  
CC STATE: WASHINGTON  
CC COUNTRY: USA  
CC ZIP: 98101  
CC  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: Apple Macintosh  
CC OPERATING SYSTEM: Apple System 7.1  
CC SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CC CURRENT APPLICATION DATA:  
CC FILING DATE: 30-MAY-1995  
CC CLASSIFICATION: 530  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/240,124  
CC FILING DATE: 09-MAY-1994  
CC APPLICATION NUMBER: US 08/161,132  
CC FILING DATE: 03-DEC-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/114,426  
CC FILING DATE: 30-AUG-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/109,745  
CC FILING DATE: 20-AUG-1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: SEESE, KATHRYN A.  
CC REGISTRATION NUMBER: 32,172  
CC REFERENCE/DOCKET NUMBER: 2814-C  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (206) 587-0430  
CC TELEFAX: (206) 233-0644  
CC TELEX: 756822  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 238 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 238 AA; 26350 MW; 315668 CN;  
Query Match 9.3%; Score 84; DB 1; Length 238;  
Best Local Similarity 68.4%; Pred. No. 2.57e+01;  
Matches 13; Conservative 2; Mismatches 3; Indels 1; Gaps 1;  
Db 2 AAAPLLLLLLVPVPLPL 20  
|| ||||| | :|||  
QY 4 AAVTLLLLLLPP-ALLSL 21  
RESULT 5  
ID US-08-240-124-2 STANDARD; PRT; 238 AA.  
XX  
AC xxxxxx  
DT  
DE  
Sequence 2, Application US/08240124  
Sequence 2, Application US/08240124  
Patent No. 551658  
GENERAL INFORMATION:  
APPLICANT: BECKMANN, M. P.  
APPLICANT: CERRETTI, DOUGLAS P.  
TITLE OF INVENTION: CYTOKINE THAT BINDS THE CELL SURFACE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:

CC ADDRESSEE: IMMUNEX CORPORATION  
CC STREET: 51 UNIVERSITY STREET  
CC CITY: SEATTLE  
CC STATE: WASHINGTON  
CC COUNTRY: USA  
CC ZIP: 98101  
CC  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: Apple Macintosh  
CC OPERATING SYSTEM: Apple System 7.1  
CC SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/240,124  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/161,132  
CC FILING DATE: 03-DEC-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/114,426  
CC FILING DATE: 30-AUG-1993  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/109,745  
CC FILING DATE: 20-AUG-1993  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: SEESE, KATHRYN A.  
CC REGISTRATION NUMBER: 32,172  
CC REFERENCE/DOCKET NUMBER: 2814-C  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (206) 587-0430  
CC TELEFAX: (206) 233-0644  
CC TELEX: 756822  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 238 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 238 AA; 26350 MW; 315668 CN;  
Query Match 9.3%; Score 84; DB 1; Length 238;  
Best Local Similarity 68.4%; Pred. No. 2.57e+01;  
Matches 13; Conservative 2; Mismatches 3; Indels 1; Gaps 1;  
Db 2 AAAPLLLLLLVPVPLPL 20  
|| ||||| | :|||  
QY 4 AAVTLLLLLLPP-ALLSL 21  
RESULT 6  
ID US-08-026-138E-3 STANDARD; PRT; 1239 AA.  
XX  
AC xxxxxx  
DT  
DE  
Sequence 3, Application US/08026138E  
Sequence 3, Application US/08026138E  
Patent No. 5502166  
GENERAL INFORMATION:  
APPLICANT: Masayoshi MISHINA  
TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nishiohata Residence 1-107  
STREET: 5214, Nishiohata-machi  
CITY: Niigata-shi  
STATE: Niigata-ken  
COUNTRY: JAPAN  
ZIP: 951  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage



CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/243,545  
CC FILING DATE: 11-MAY-1994  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US 08/162,407  
CC FILING DATE: 03-DEC-1993  
CC APPLICATION NUMBER: 08/111,758  
CC FILING DATE: August 25, 1993  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/106,463  
CC FILING DATE: August 12, 1993  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/068,394  
CC FILING DATE: May 24, 1993  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Malaska, Stephen L.  
CC REGISTRATION NUMBER: 32,655  
CC REFERENCE/DOCKET NUMBER: 2813-C  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (206) 587-0430  
CC TELEFAX: (206) 233-0644  
CC TELEX: 756822  
CC INFORMATION FOR SEQ ID NO: 6:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 235 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 235 AA; 26415 MW; 293990 CN;  
  
Query Match 9.2%; Score 83; DB 1; Length 235;  
Best Local Similarity 55.6%; Pred. No. 3.02e+01;  
Matches 10; Conservative 6; Mismatches 1; Indels 1; Gaps 1;  
  
Db 7 AWSPTVYLLLLSSGL 24  
QY 2 PWAAVT-LLLLLLLPPAL 18  
  
RESULT 9  
ID PCT-US94-05365-6 STANDARD; PRT; 235 AA.  
XX  
AC xxxxxx  
XX  
DT  
DE  
Sequence 6, Application PC/TUS9405365  
Sequence 6, Application PC/TUS9405365  
GENERAL INFORMATION:  
APPLICANT: Lyman, Stewart D.  
APPLICANT: Beckmann, M. Patricia  
TITLE OF INVENTION: Ligands for flt3/flk-2 Receptors  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Stephen L. Malaska, Immunex Corporation  
STREET: 51 University Street  
City: Seattle  
State: Washington  
COUNTRY: US  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/05365  
FILING DATE: May 24, 1994

CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: -to be assigned-  
CC FILING DATE: May 11, 1994  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/209,502  
CC FILING DATE: March 7, 1994  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/162,407  
CC FILING DATE: December 3, 1993  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/111,758  
CC FILING DATE: August 25, 1993  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/106,463  
CC FILING DATE: August 12, 1993  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/068,394  
CC FILING DATE: May 24, 1993  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Malaska, Stephen L.  
CC REGISTRATION NUMBER: 32,655  
CC REFERENCE/DOCKET NUMBER: 2813-B  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (206) 587-0430  
CC TELEFAX: (206) 233-0644  
CC TELEX: 756822  
CC INFORMATION FOR SEQ ID NO: 6:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 235 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 235 AA; 26415 MW; 293990 CN;  
  
Query Match 9.2%; Score 83; DB 3; Length 235;  
Best Local Similarity 55.6%; Pred. No. 3.02e+01;  
Matches 10; Conservative 6; Mismatches 1; Indels 1; Gaps 1;  
  
Db 7 AWSPTVYLLLLSSGL 24  
QY 2 PWAAVT-LLLLLLLPPAL 18  
  
RESULT 10  
ID PCT-US91-06532-2 STANDARD; PRT; 263 AA.  
XX  
AC xxxxxx  
XX  
DT  
DE  
Sequence 2, Application PC/TUS9106532  
Sequence 2, Application PC/TUS9106532  
GENERAL INFORMATION:  
APPLICANT: Roizman, Bernard  
TITLE OF INVENTION: Recombinant Herpes Simplex Viruses  
TITLE OF INVENTION: Vaccines and Methods  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
ADDRESS: Bicknell  
STREET: Two First National Plaza Suite 2100  
City: Chicago  
State: Illinois  
COUNTRY: USA  
ZIP: 60603



\*\*\*\*\*

WQSEFH  
(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:16:51 1999; MasPar time 5.77 Seconds  
Tabular output not generated. 602.480 Million cell updates/sec

Title: >US-08-938-548B-10  
Description: (1-123) from US089938548B.pep  
Perfect Score: 899  
Sequence: 1 VPWAAVTLTLLLLPPALLS.....GRGCPVTVTALAPRGSGV 123

Scoring table: PAM 150  
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot37  
1:Swissprot

Statistics: Mean 41.468; Variance 75.472; scale 0.549

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	103	11.5	567	1	PLATELET GLYCOPROTEIN	8.73e-03
2	101	11.2	131	1	SECRETIN PRECURSOR (FR	1.73e-02
3	96	10.7	566	1	FBLA HUMAN	9.21e-02
4	96	10.7	601	1	FBLB HUMAN	9.21e-02
5	96	10.7	683	1	FBLD HUMAN	9.21e-02
6	96	10.7	703	1	FBLG HUMAN	9.21e-02
7	91	10.1	440	1	LCAT HUMAN	4.64e-01
8	89	9.9	602	1	PGH1 RAT	8.72e-01
9	89	9.9	1061	1	ANPA HUMAN	8.72e-01
10	88	9.8	234	1	4-1BB LIGAND (4-1BBL)	1.19e+00
11	88	9.8	438	1	LCAT MOUSE	1.19e+00
12	86	9.6	322	1	RLUC HAEIN	2.20e+00
13	86	9.6	440	1	LCAT PAPAN	2.20e+00
14	86	9.6	497	1	SC14 YARLI	2.20e+00
15	86	9.6	1663	1	COMPLEMENT C3 PRECURSO	2.20e+00
16	85	9.5	236	1	PLACENTAL CTG	2.98e+00
17	85	9.5	317	1	LIP1 PSYIM	2.98e+00
18	85	9.5	334	1	FEPD ECOLI	2.98e+00
19	85	9.5	1027	1	CAFE RIFPA	2.98e+00
20	84	9.3	90	1	LYSIS PROTEIN (E PROTE	4.03e+00
21	84	9.3	90	1	VGE_BPS13	4.03e+00
22	84	9.3	238	1	EFA3 HUMAN	4.03e+00
23	84	9.3	251	1	C1QB_HUMAN	4.03e+00

24	84	9.3	315	1	LIP3_MORSP	LIPASE 3 PRECURSOR (EC	4.03e+00
25	84	9.3	319	1	RLUC_ECOLI	RIBOSOMAL LARGE SUBUNI	4.03e+00
26	84	9.3	384	1	RN_DROME	GTPASE ACTIVATING PROT	4.03e+00
27	84	9.3	696	1	LSHR_PIG	LUTROPIN-CHORIOGONADOT	4.03e+00
28	84	9.3	1001	1	PTPX_MOUSE	PROTEIN-TYROSINE PHOSP	4.03e+00
29	84	9.3	2541	1	TALI_MOUSE	TALIN	4.03e+00
30	83	9.2	208	1	GPBB_PAPCY	PLATELET GLYCOPROTEIN	5.43e+00
31	83	9.2	224	1	OXO2_HORVU	OXALATE OXIDASE PRECUR	5.43e+00
32	83	9.2	235	1	FL3L_HUMAN	SL CYTOKINE PRECURSOR	5.43e+00
33	83	9.2	248	1	ICP3_HSV11	INFECTED CELL PROTEIN	5.43e+00
34	83	9.2	252	1	ICP3_HSV1D	INFECTED CELL PROTEIN	5.43e+00
35	83	9.2	263	1	ICP3_HSV1F	INFECTED CELL PROTEIN	5.43e+00
36	83	9.2	440	1	LCAT_RABIT	PHOSPHATIDYLCHOLINE-ST	5.43e+00
37	83	9.2	440	1	LCAT_RAT	PHOSPHATIDYLCHOLINE-ST	5.43e+00
38	83	9.2	536	1	YABK_ECOLI	HYPOTHETICAL 59.6 KD P	5.43e+00
39	83	9.2	615	1	ALBU_CHICK	SERUM ALBUMIN PRECURSO	5.43e+00
40	83	9.2	628	1	FTSH_PORPU	CELL DIVISION PROTEIN	5.43e+00
41	83	9.2	676	1	ICP0_HSVBK	TRANS-ACTING TRANSCRIP	5.43e+00
42	83	9.2	676	1	ICP0_HSVBJ	TRANS-ACTING TRANSCRIP	5.43e+00
43	83	9.2	1004	1	PTPX_RAT	PROTEIN-TYROSINE PHOSP	5.43e+00
44	83	9.2	1109	1	CYGD_CANFA	RETINAL GUANYLYL CYCLA	5.43e+00
45	83	9.2	1310	1	ACE_RABIT	ANGIOTENSIN-CONVERTING	5.43e+00

ALIGNMENTS

RESULT 1  
ID GPV\_RAT STANDARD; PRT; 567 AA.  
AC 008770;  
DT 15-JUL-1998 (REL. 36, CREATED)  
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE PLATELET GLYCOPROTEIN V PRECURSOR (GPV) (CD42D).  
GN GP5.  
OS RATTUS NORVEGICUS (RAT).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WISTAR; TISSUE=LIVER;  
RX MEDLINE: 97275136.  
RA RAVANAT C., MORALES M., AZORSA D.O., MOOG S., SCHUHLER S.,  
RA GRUNERT P., LOW D., VAN DORSELAER A., CAZENAVE J.-P., LANZA F.;  
RT "Gene cloning of rat and mouse platelet glycoprotein V;  
RT identification of megakaryocyte-specific promoters and demonstration  
RT of functional thrombin cleavage.";  
RL BLOOD 89:3253-3262(1997).  
CC -!- FUNCTION: THE GPIIb-V-III COMPLEX FUNCTIONS AS THE VON WILLEBRAND  
CC FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT  
CC PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO  
CC INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A  
CC CRITICAL INITIATING EVENT IN HEMOSTASIS (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN  
CC MANY PROTEINS. NUMBER IN THIS PROTEIN: 15.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: Z69594; E222201; -  
CC PFAM: PF00560; LRR; 8.  
CC PLATELET; TRANSMEMBRANE; GLYCOPROTEIN; BLOOD COAGULATION;  
KW REPEAT; LEUCINE-REPEAT; CELL ADHESION; SIGNAL.  
FT SIGNAL 1 16 POTENTIAL.  
FT CHAIN 17 567 PLATELET GLYCOPROTEIN V.  
FT DOMAIN 17 522 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 523 543 POTENTIAL.  
FT DOMAIN 544 567 CYTOPLASMIC (POTENTIAL).

```

QY 7 TLLLLLLPPALLSLGVDAQPLP 29
RESULT 3
ID FBLA_HUMAN STANDARD; PRT; 566 AA.
AC P23142;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE FIBULIN-1, ISOFORM A PRECURSOR.
GN FBLN1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91100426.
RA AGRANES W.S., TRAN H., BURGESS W.H., DICKERSON K.;
RT "Fibulin is an extracellular matrix and plasma glycoprotein with
RT repeated domain structure."
RL J. CELL BIOL. 111:3155-3164(1990).
RN [2]
RP SEQUENCE OF 30-44.
RX MEDLINE; 89354537.
RA AGRANES W.S., DICKERSON K., BURGESS W.H., RUOSLAHTI E.;
RT "Fibulin, a novel protein that interacts with the fibronectin
RT receptor beta subunit cytoplasmic domain."
RL CELL 58:623-629(1989).
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC (AC P23143), C (AC P23144) AND D (AC P37888); DIFFERING ONLY IN
CC THEIR C-TERMINAL REGIONS, ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: CONTAINS 3 ANAPHYLATOXIN-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 8 COMPLETE AND ONE INCOMPLETE EGF-LIKE
CC DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53741; G31415; -
DR PIR; A32826; A32826.
DR PIR; A36346; A36346.
DR MIM; 135820; -
DR PROSITE; PS00010; ASX_HYDROXYL; 4.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01177; ANAPHYLATOXIN_1; 3.
DR PROSITE; PS01178; ANAPHYLATOXIN_2; 3.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01187; EGF_CA; 8.
DR PFAM; PF00008; EGF; 5.
DR HSPF; P35555; IEMO.
DR SIGNAL; ALTERNATIVE SPLICING; GLYCOPROTEIN; EXTRACELLULAR MATRIX;
KW REPEAT; PLASMA; EGF-LIKE DOMAIN; CALCIUM-BINDING.
FT SIGNAL 1 29
FT CHAIN 30 566
FT DOMAIN 36 144
FT REPEAT 36 76
FT REPEAT 77 111
FT REPEAT 112 144
FT REPEAT 176 215
FT DOMAIN 216 261
FT DOMAIN 262 307
FT DOMAIN 308 355
FT DOMAIN 356 398
FT DOMAIN 399 440
FT DOMAIN 441 480
FT DOMAIN 481 524
FT DOMAIN 525 566

```

```

FT DISULFID 36 61
FT DISULFID 37 68
FT DISULFID 50 69
FT DISULFID 78 109
FT DISULFID 91 110
FT DISULFID 112 136
FT DISULFID 113 143
FT DISULFID 126 144
FT DISULFID 180 190
FT DISULFID 186 199
FT DISULFID 201 214
FT DISULFID 220 233
FT DISULFID 227 242
FT DISULFID 248 260
FT DISULFID 266 279
FT DISULFID 273 288
FT DISULFID 294 306
FT DISULFID 312 325
FT DISULFID 319 334
FT DISULFID 341 354
FT DISULFID 360 373
FT DISULFID 367 382
FT DISULFID 384 397
FT DISULFID 403 415
FT DISULFID 411 424
FT DISULFID 426 439
FT DISULFID 445 454
FT DISULFID 450 463
FT DISULFID 465 479
FT DISULFID 485 498
FT DISULFID 494 507
FT DISULFID 509 523
FT CARBOHYD 98 98
FT CARBOHYD 447 447
FT CARBOHYD 535 535
FT CARBOHYD 539 539
FT CONFLICT 36 36
FT CONFLICT 41 42
SQ SEQUENCE 566 AA; 61593 MW; A2D23E14 CRC32;
Query Match 10.7%; Score 96; DB 1; Length 566;
Best Local Similarity 57.1%; Pred. No. 9.21e-02;
Matches 16; Conservative 3; Mismatches 8; Indels 1; Gaps 1;
Db 10 VPLPLLLGLLALLAAGVDADVLEACC 37
| | | | | | | | | | | | | | | | | |
QY 6 VTLILLLLPPALLSLGVDAQPLPD-CC 32
RESULT 4
ID FBLB_HUMAN STANDARD; PRT; 601 AA.
AC P23143;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE FIBULIN-1, ISOFORM B PRECURSOR.
GN FBLN1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91100426.
RA AGRANES W.S., TRAN H., BURGESS W.H., DICKERSON K.;
RT "Fibulin is an extracellular matrix and plasma glycoprotein with
RT repeated domain structure."
RL J. CELL BIOL. 111:3155-3164(1990).
RN [2]
RP SEQUENCE OF 30-44.
RX MEDLINE; 89354537.
RA AGRANES W.S., DICKERSON K., BURGESS W.H., RUOSLAHTI E.;
RT "Fibulin, a novel protein that interacts with the fibronectin
RT receptor beta subunit cytoplasmic domain."

```

[illegible]

RA FUNKE H., VON ECKARDSTEIN A., PRITCHARD P.H., HORNEY A.E.,  
 RA WIEBUSCH H., MOTTI C., HAYDEN M.R., DACHET C., JACOTOT B., GERDES U.,  
 RA FAERGEMAN O., ALBERS J.J., COLLEONI N., CATAPANO A., FROHLICH J.,  
 RA ASSMANN G.;  
 RT "Genetic and phenotypic heterogeneity in familial lecithin:  
 RT cholesterol acyltransferase (LCAT) deficiency. Six newly identified  
 RT defective alleles further contribute to the structural heterogeneity  
 RT in this disease.";  
 RT J. CLIN. INVEST. 91:677-683(1993).  
 RL [14]  
 RN VARIANTS THR-117 AND CYS-182.  
 RX MEDLINE; 93305754.  
 RP HILL J.S., O.K., WANG X., PRITCHARD P.H.;  
 RT "Lecithin:cholesterol acyltransferase deficiency: identification of a  
 RT causative gene mutation and a co-inherited protein polymorphism.";  
 RL BIOCHIM. BIOPHYS. ACTA 1181:321-323(1993).  
 RN [15]  
 RN VARIANT HIS-164.  
 RX MEDLINE; 95331753.  
 RP STEYRER E., HAUBENWALLNER S., HOERL G., GIESHAUF W., KOSTNER G.M.,  
 RA ZECHNER R.;  
 RT "A single G to A nucleotide transition in exon IV of the lecithin:  
 RT cholesterol acyltransferase (LCAT) gene results in an Arg140 to His  
 RT substitution and causes LCAT-deficiency.";  
 RL HUM. GENET. 96:105-109(1995).  
 RN [16]  
 RN VARIANTS ARG-57 AND 10 AA INSERTION IN POSITION 17.  
 RP MEDLINE; 95227171.  
 RX WIEBUSCH H., CULLEN P., OWEN J.S., COLLINS D., SHARP P.S., FUNKE H.,  
 RA ASSMANN G.;  
 RT "Deficiency of lecithin:cholesterol acyltransferase due to compound  
 RT heterozygosity of two novel mutations (Gly33Arg and 30 bp ins) in the  
 RT LCAT gene.";  
 RL HUM. MOL. GENET. 4:143-145(1995).  
 RN [17]  
 RN VARIANT SER-54.  
 RX MEDLINE; 96400966.  
 RP OWEN J.S., WIEBUSCH H., CULLEN P., WATTS G.F., LIMA V.L.M.,  
 RA FUNKE H., ASSMANN G.;  
 RT "Complete deficiency of plasma lecithin-cholesterol acyltransferase  
 RT (LCAT) activity due to a novel homozygous mutation (Gly-30-Ser) in  
 RT the LCAT gene.";  
 RL HUM. MUTAT. 8:79-82(1996).  
 CC -!- FUNCTION: CENTRAL ENZYME IN THE EXTRACELLULAR METABOLISM OF PLASMA  
 CC LIPOPROTEINS. AMONG OTHER SUBSTRATES IT ESTERIFIES THE FREE  
 CC CHOLESTEROL TRANSPORTED IN PLASMA LIPOPROTEINS.  
 CC -!- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + STEROL = STEROL ESTER +  
 CC 1-ACYLGLYCEROPHOSPHOCHOLINE (PALMITOYL, OLEOYL, AND LINOLEOYL CAN  
 CC BE TRANSFERRED; A NUMBER OF STEROLS, INCLUDING CHOLESTEROL, CAN  
 CC ACT AS ACCEPTOR).  
 CC -!- ENZYME REGULATION: APOLIPOPROTEIN A-I IS A POTENT ACTIVATOR FOR  
 CC THIS ENZYME.  
 CC -!- DISEASE: DEFECTS IN LCAT ARE THE CAUSE OF NORUM AND FISH EYE  
 CC DISEASES.  
 CC -!- SIMILARITY: PARTIAL WITH LIPASES (PANCREATIC, GASTRIC, HEPATIC,  
 CC LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M12625; G307117; -;  
 CC EMBL; X06537; -; NOT ANNOTATED\_CDS.  
 CC EMBL; M26268; G187025; -;  
 CC EMBL; X04981; G34287; -;  
 CC EMBL; M17959; G386836; -;  
 CC PIR; A00571; XXHUN.  
 CC PIR; A29661; A29661.  
 CC PIR; A25575; A25575.

DR PIR; JQ0036; JQ0036.  
 DR MIM; 136120; -;  
 DR MIM; 245900; -;  
 DR PROSITE; PS00120; LIPASE\_SER; 1.  
 KW TRANSFERASE; ACYLTRANSFERASE; LIPID METABOLISM; GLYCOPROTEIN; SIGNAL;  
 KW POLYMORPHISM; DISEASE MUTATION.  
 FT SIGNAL 1 24  
 FT CHAIN 25 440 LECITHIN-CHOLESTEROL ACYLTRANSFERASE.  
 FT ACT\_SITE 205 205 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 74 98  
 FT DISULFID 337 380  
 FT CARBOHYD 44 44  
 FT CARBOHYD 108 108  
 FT CARBOHYD 296 296  
 ...  
 Note: remainder of annotations omitted.  
 Query Match 10.18; Score 91; DB 1; Length 440;  
 Best Local Similarity 81.38; Pred. No. 4.64e-01;  
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Db 7 PQQWVTLGLGLLPPA 22  
 QY 2 PWAAYVTLGLLGLLPPA 17  
 RESULT 8  
 ID PGH1\_RAT STANDARD; PRT; 602 AA.  
 AC Q63971; Q62731; Q63684;  
 DT 15-DEC-1998 (REL. 37, CREATED)  
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)  
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
 DE PROSTAGLANDIN G/H SYNTHASE 1 PRECURSOR (EC 1.14.99.1) (CYCLOOXYGENASE  
 DE -1) (COX-1) (PROSTAGLANDIN-ENDOPEROXIDE SYNTHASE 1) (PROSTAGLANDIN H2  
 DE SYNTHASE 1) (PGH SYNTHASE 1) (PGHS-1) (PHS 1).  
 GN PTGSI OR COXI OR COX-1.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY;  
 RX MEDLINE; 94099619.  
 RA FENG L., SUN Y., XIA Y., TANG W.W., CHANMUGAM P., SOYOOLA E.,  
 RA WILSON C.B., HWANG D.;  
 RT "Cloning two isoforms of rat cyclooxygenase: differential regulation  
 RT of their expression.";  
 RL ARCH. BIOCHEM. BIOPHYS. 307:361-368(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER 344; TISSUE=TRACHEA;  
 RX MEDLINE; 95168876.  
 RA KITZLER J., HILL E., HARDMAN R., REDDY N., PHILPOT R., ELING T.E.;  
 RT "Analysis and quantitation of splicing variants of the TPA-inducible  
 RT PGHS-1 mRNA in rat tracheal epithelial cells.";  
 RL ARCH. BIOCHEM. BIOPHYS. 316:856-863(1995).  
 CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN REGULATING OR PROMOTING  
 CC CELL PROLIFERATION IN SOME NORMAL AND NEOPLASTICALLY TRANSFORMED  
 CC CELLS.  
 CC -!- CATALYTIC ACTIVITY: ARACHIDONATE + AH(2) + 2 O(2) = PROSTAGLANDIN  
 CC H2 + A + H(2)O.  
 CC -!- PATHWAY: FIRST STEP IN THE FORMATION OF PROSTAGLANDINS AND  
 CC THROMBOXANES.  
 CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MICROSOMAL MEMBRANE.  
 CC -!- THIS ENZYME ACTS BOTH AS A DIOXYGENASE AND AS A PEROXIDASE.  
 CC -!- THIS ENZYME IS THE TARGET OF NONSTEROIDAL ANTI-INFLAMMATORY DRUGS  
 CC SUCH AS ASPIRIN.  
 CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -!- SIMILARITY: BELONGS TO THE PROSTAGLANDIN G/H SYNTHASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

FT SIGNAL 1 32  
 FT CHAIN 33 1061  
 FT DOMAIN 33 473  
 FT TRANSMEM 474 494  
 FT DOMAIN 495 1061  
 FT DOMAIN 528 805  
 FT DISULFID 92 118  
 FT DISULFID 196 245  
 FT DISULFID 455 455  
 FT DISULFID 464 464  
 FT CARBOHYD 34 34  
 FT CARBOHYD 45 45  
 FT CARBOHYD 212 212  
 FT CARBOHYD 338 338  
 FT CARBOHYD 379 379  
 FT CARBOHYD 386 386  
 FT CARBOHYD 427 427  
 SQ SEQUENCE 1061 AA; 118918 MW; C6117B45 CRC32;

Query Match 9.8%; Score 89; DB 1; Length 1061;  
 Best Local Similarity 68.2%; Pred. No. 8.72e-01;  
 Matches 15; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Db 14 LLLLLLPPPLLLLRGSHAGN 35  
 QY 8 LLLLLLPPALLSL-GVDAQPL 28

RESULT 10  
 ID 41BL.HUMAN STANDARD; PRT; 254 AA.  
 AC P41273; 1995 (REL. 31, CREATED)  
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE 4-LBB LIGAND (4-1BBL).  
 OS HOMO SAPIENS (HUMAN).  
 OC EKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94374434.  
 RA ALDERSON M.R., SMITH C.A., TOUGH T.W., DAVIS-SMITH T., ARMITAGE R.J.,  
 RA FALK B., ROUX E., BAKER E., SUTHERLAND G.R., DIN W.S., GOODWIN R.G.;  
 RT "Molecular and biological characterization of human 4-lbb and its  
 RT ligand."  
 RL EUR. J. IMMUNOL. 24:2219-2227(1994).  
 CC -!- FUNCTION: INDUCES THE PROLIFERATION OF ACTIVATED PERIPHERAL BLOOD  
 CC T CELLS. MAY HAVE A ROLE IN ACTIVATION-INDUCED CELL DEATH (AICD).  
 CC MAY PLAY A ROLE IN COGNATE INTERACTIONS BETWEEN T CELLS AND  
 CC B CELLS/MACROPHAGES.  
 CC -!- SUBUNIT: HOMOTRIMER (POTENTIAL).  
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, PLACENTA, LUNG, SKELETAL  
 CC MUSCLE AND KIDNEY.  
 CC -!- SIMILARITY: BELONGS TO THE TUMOR NECROSIS FACTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U03398; G571323; -  
 CC PROSITE; PS00251; TNF\_1; 1.  
 CC PROSITE; PS00049; TNF\_2; 1.  
 CC PFAM; PF00229; TNF; 1  
 CC CYTOKINE; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL-ANCHOR.  
 FT DOMAIN 1 28  
 FT TRANSMEM 29 49  
 FT DOMAIN 50 254  
 FT DOMAIN 35 41

SQ SEQUENCE 254 AA; 26624 MW; C68C1B27 CRC32;

Query Match 9.8%; Score 88; DB 1; Length 254;  
 Best Local Similarity 66.7%; Pred. No. 1.19e+00;  
 Matches 12; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

Db 27 LPWALVAGLLLLLLAA 44  
 QY 1 VPWAAVT-LLLLLLPPA 17

RESULT 11  
 ID LCAT\_MOUSE STANDARD; PRT; 438 AA.  
 AC P16301;  
 DT 01-AUG-1990 (REL. 15, CREATED)  
 DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE PHOSPHATIDYLCHOLINE-STEROL ACYLTRANSFERASE PRECURSOR (EC 2.3.1.43)  
 DE (LECITHIN-CHOLESTEROL ACYLTRANSFERASE) (PHOSPHOLIPID-CHOLESTEROL  
 DE ACYLTRANSFERASE).  
 DE LCAT.  
 GN MUS MUSCULUS (MOUSE).  
 OS MUS MUSCULUS (MOUSE).  
 OC EKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 90094326.  
 RA WARDEN C.H., LANGNER C.A., GORDON J.I., TAYLOR B.A., MCLEAN J.W.,  
 RA LUSIS A.J.;  
 RT "Tissue-specific expression, developmental regulation, and  
 RT chromosomal mapping of the lecithin: cholesterol acyltransferase  
 RT gene. Evidence for expression in brain and testes as well as liver."  
 RL J. BIOL. CHEM. 264:21573-21581(1989).  
 RN [2]  
 RP SEQUENCE OF 1-14 FROM N.A.  
 RA MERONI G., MARGARETTI N., MAGNAGHI P., TARMELLI R.;  
 RL SUBMITTED (MAY-1992) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -!- FUNCTION: CENTRAL ENZYME IN THE EXTRACELLULAR METABOLISM OF PLASMA  
 CC LIPOPROTEINS. AMONG OTHER SUBSTRATES IT ESTERIFIES THE FREE  
 CC CHOLESTEROL TRANSPORTED IN PLASMA LIPOPROTEINS.  
 CC -!- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + STEROL = STEROL ESTER +  
 CC 1-ACYLGLYCEROPHOSPHOCHOLINE (PALMITOYL, OLEOYL, AND LINOLEOYL CAN  
 CC BE TRANSFERRED; A NUMBER OF STEROLS, INCLUDING CHOLESTEROL, CAN  
 CC ACT AS ACCEPTOR).  
 CC -!- ENZYME REGULATION: APOLIPOPROTEIN A-I IS A POTENT ACTIVATOR FOR  
 CC THIS ENZYME.  
 CC -!- SIMILARITY: PARTIAL WITH LIPASES (PANCREATIC, GASTRIC, HEPATIC,  
 CC LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; J05154; G293697; -  
 CC EMBL; X54095; G52874; -  
 CC PIR; A34158; XXMSN.  
 CC MGD; MGI:96755; LCAT.  
 CC PROSITE; PS00120; LIPASE\_LSER; 1.  
 CC TRANSFERASE; ACYLTRANSFERASE; LIPID METABOLISM; GLYCOPROTEIN; SIGNAL.  
 FT SIGNAL 1 24  
 FT CHAIN 25 438  
 FT ACT\_SITE 205 205  
 FT DISULFID 74 98  
 FT DISULFID 337 380  
 FT CARBOHYD 44 44  
 FT CARBOHYD 108 108  
 FT CARBOHYD 296 296  
 FT CARBOHYD 397 397  
 FT CARBOHYD 408 408

OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYCETES; SACCHAROMYCETALES;  
 OC DIPODASCACAE; YARROWIA.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 20460 / W29;  
 RX MEDLINE; 94186533.  
 RA LOPEZ M.C., NICAUD J.-M., SKINNER H.B., VERGNOLLE C., KADER J.-C.,  
 RA BANKAITIS V., GAILLARDIN C.;  
 RT "A phosphatidylinositol/phosphatidylcholine transfer protein is  
 RT required for differentiation of the dimorphic yeast *Yarrowia*  
 RT *lipolytica* from the yeast to the mycelial form.";  
 RL J. CELL BIOL. 124:113-127(1994).  
 CC -!- FUNCTION: REQUIRED FOR TRANSPORT OF SECRETORY PROTEINS FROM THE  
 CC GOLGI COMPLEX. CATALYZES THE TRANSFER OF PHOSPHATIDYLINOSITOL AND  
 CC PHOSPHATIDYLCHOLINE BETWEEN MEMBRANES IN VITRO.  
 CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE GOLGI COMPLEX AS A  
 CC PERIPHERAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE SEC14 CYTOSOLIC FACTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L20972; G311167; -  
 DR PFAM; PF00650; CRAL\_Trio; 1.  
 KW HSSP; P24280; 1AUA.  
 KW TRANSPORT; PROTEIN STACK.  
 FT DOMAIN 300 320 PRO-RICH.  
 FT DOMAIN 328 372 LEU-RICH.  
 SQ SEQUENCE 497 AA; 55982 MW; 6209E8FC CRC32;  
 -----  
 Query Match 9.6%; Score 86; DB 1; Length 497;  
 Best Local Similarity 45.0%; Pred. No. 2.20e+00;  
 Matches 18; Conservative 9; Mismatches 8; Indels 5; Gaps 5;  
 -----  
 Db 351 AVLLLELLPSRLPLPRQGRSRCCQDRHCPC 390  
 QY 5 AVTLILL-LLLPALL-S-LGVDA-Q-PLPDCRCQKTCSC 39  
 -----  
 RESULT 15  
 ID CO3\_RAT STANDARD; PRT; 1663 AA.  
 AC P01026;  
 DT 21-JUL-1986 (REL. 01, CREATED)  
 DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
 DE COMPLEMENT C3 PRECURSOR [CONTAINS: C3A ANAPHYLATOXIN].  
 GN C3.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.  
 RN [1]  
 RC STRAIN=WISTAR; TISSUE=LIVER;  
 RP MEDLINE; 90245672.  
 RX MISUMI Y., SOHDA M., IKEHARA Y.;  
 RA "Nucleotide and deduced amino acid sequence of rat complement C3.";  
 RT NUCLEIC ACIDS RES. 18:2178-2178(1990).  
 RL [2]  
 RN SEQUENCE OF 671-748.  
 RP MEDLINE; 79062262.  
 RX JACOBS J.W., RUBIN J.S., HUGLI T.E., BOGARDT R.A., MARIZ I.K.,  
 RA DANIELS J.S., DAUGHADAY W.H., BRADSHAW R.A.;  
 RT "Purification, characterization, and amino acid sequence of rat  
 RT anaphylatoxin (C3a)."  
 RL BIOCHEMISTRY 17:5031-5038(1978).  
 RN [3]  
 RP SEQUENCE OF 1316-1595 FROM N.A.  
 RX MEDLINE; 89380332.  
 -----  
 RA SUNDSTROM S.A., KOMM B.S., PONCE-DE-LEON H., YI Z., TEUSCHER C.,  
 RA LYTLE C.R.;  
 RT "Estrogen regulation of tissue-specific expression of complement C3.";  
 RL J. BIOL. CHEM. 264:16941-16947(1989).  
 CC -!- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE  
 CC COMPLEMENT SYSTEM. ITS PROCESSING BY C3 CONVERTASE IS THE CENTRAL  
 CC REACTION IN BOTH CLASSICAL AND ALTERNATIVE COMPLEMENT PATHWAYS.  
 CC AFTER ACTIVATION C3B CAN BIND COVALENTLY, VIA ITS REACTIVE  
 CC THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR IMMUNE AGGREGATES.  
 CC -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C3,  
 CC C3A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT  
 CC INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR  
 CC PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND  
 CC BASOPHILIC LEUKOCYTES.  
 CC -!- SUBUNIT: C3 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 ARG  
 CC RESIDUES, FORMING TWO CHAINS, BETA' & ALPHA', LINKED BY A DISULFIDE  
 CC BOND. C3 CONVERTASE ACTIVATES C3 BY CLEAVING THE ALPHA CHAIN.  
 CC RELEASING C3A ANAPHYLATOXIN & GENERATING C3B (BETA CHAIN + ALPHA'  
 CC CHAIN).  
 CC -!- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.  
 CC -!- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X52477; G56954; -  
 DR EMBL; M29866; G554423; ALT\_SEQ.  
 DR PIR; A01260; A01260.  
 DR PIR; S15764; S15764.  
 DR PROSITE; PS00477; ALPHA\_2\_MACROGLOBULIN; 1.  
 DR PROSITE; PS01177; ANAPHYLATOXIN\_1; 1.  
 DR PROSITE; PS01178; ANAPHYLATOXIN\_2; 1.  
 DR PFAM; PF00207; A2M; 1.  
 DR HSSP; P01024; 1C3D.  
 KW COMPLEMENT PATHWAY; COMPLEMENT ALTERNATE PATHWAY; PLASMA;  
 KW INFLAMMATORY RESPONSE; GLYCOPROTEIN; SIGNAL.  
 FT SIGNAL 1 24  
 FT CHAIN 25 1663 COMPLEMENT C3.  
 FT CHAIN 25 666 BETA CHAIN.  
 FT CHAIN 671 1663 ALPHA CHAIN.  
 FT PEPTIDE 671 748 C3A ANAPHYLATOXIN.  
 FT CHAIN 749 1663 C3B (ALPHA' CHAIN).  
 FT SITE 748 749 CLEAVAGE (BY C3 CONVERTASE).  
 FT DOMAIN 693 728 ANAPHYLATOXIN-LIKE.  
 FT DISULFID 626 661 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 693 720 BY SIMILARITY.  
 FT DISULFID 694 727 BY SIMILARITY.  
 FT DISULFID 707 728 BY SIMILARITY.  
 FT DISULFID 873 1513 BY SIMILARITY.  
 FT DISULFID 1101 1158 BY SIMILARITY.  
 FT DISULFID 1358 1489 BY SIMILARITY.  
 FT DISULFID 1389 1458 BY SIMILARITY.  
 FT DISULFID 1506 1511 BY SIMILARITY.  
 FT DISULFID 1518 1590 BY SIMILARITY.  
 FT DISULFID 1537 1661 BY SIMILARITY.  
 FT THIOLEST 1010 1013 PROBABLE.  
 FT CARBOHYD 939 939 PROBABLE.  
 FT CARBOHYD 1617 1617 PROBABLE.  
 FT CONFLICT 721 722 LK -> KL (IN REF. 2).  
 SQ SEQUENCE 1663 AA; 186460 MW; 0428CF63 CRC32;  
 -----  
 Query Match 9.6%; Score 86; DB 1; Length 1663;  
 Best Local Similarity 66.7%; Pred. No. 2.20e+00;  
 Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 -----  
 Db 10 LVLLLLLASSLLAIG 24  
 |:|||||:|:|:|



\*\*\*\*\*

W P E R E H

(TM)

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:17:24 1999; MasPar time 11.47 Seconds  
585.113 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-938-548B-10  
Description: (1-123) from US08938548B.pep  
Perfect Score: 899  
Sequence: 1 VPWAAVTLTLLLLLPALLS.....GRGPTVTTTALAPRGSGV 123

Scoring table: PAM 150  
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrmb19

1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 39.562; Variance 86.129; scale 0.460

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	899	100.0	130	11	HYPOCRETIN (PREPRO-ORE PREPRO-OREXIN.	6.00e-145
2	853	94.9	130	11	PREPRO-OREXIN.	3.85e-136
3	734	81.6	131	4	O43612	1.82e-113
4	723	80.4	131	6	PREPRO-OREXIN PRECURSO	2.23e-111
5	113	12.6	400	4	GDNF FAMILY RECEPTOR A	3.75e-03
6	106	11.8	679	10	HYPOTHETICAL 74.6 KD P	3.22e-02
7	98	10.9	641	11	FRIZZLED-1.	3.44e-01
8	96	10.7	626	11	FRIZZLED-1.	6.13e-01
9	94	10.5	574	11	ACYLOXYACYL HYDROLASE.	1.08e+00
10	93	10.3	331	13	LEUCINE ZIPPER WITH BA	1.44e+00
11	93	10.3	1174	6	TIGHT JUNCTION PROTEIN	1.44e+00
12	92	10.2	106	10	PROLINE- AND LEUCINE-R	1.90e+00
13	92	10.2	857	5	PAIRED BOX PROTEIN PAX	1.90e+00
14	90	10.0	290	2	L-1 METALLO-BETA-LACTA	3.32e+00
15	90	10.0	833	4	SORTILIN PRECURSOR	3.32e+00
16	89	9.9	1321	4	KIAA0634 PROTEIN (FRAG	4.36e+00
17	88	9.8	226	2	HYPOTHETICAL 23.1 KD P	5.73e+00
18	88	9.8	335	13	THYROID HORMONE INDUCE	5.73e+00
19	87	9.7	252	2	HYPOTHETICAL 25.9 KD P	7.52e+00
20	87	9.7	396	5	PUTATIVE GPI-ANCHORED	7.52e+00

21	87	9.7	583	13	Q90491	DNA BINDING PROTEIN E1	7.52e+00
22	87	9.7	1238	5	Q18780	SIMILARITY TO MOUSE SM	7.52e+00
23	86	9.6	335	2	O85011	HYPOTHETICAL 34.4 KD P	9.83e+00
24	86	9.6	410	2	P72841	HYPOTHETICAL 48.1 KD P	9.83e+00
25	86	9.6	999	10	O82432	LEUCINE-RICH RECEPTOR-NB-2.	9.83e+00
26	86	9.6	1099	11	P97527	PORCINE MEMBRANE COFAC	1.28e+01
27	85	9.5	363	6	O02839	RNA-DEPENDENT RNA POLY	1.28e+01
28	85	9.5	878	14	O83101	ANION EXCHANGER ISOFOR	1.67e+01
29	84	9.3	84	11	O60471	108 PROTEIN PRECURSOR.	1.67e+01
30	84	9.3	102	10	O43495	HYPOTHETICAL 30.9 KD P	1.67e+01
31	84	9.3	277	2	O33285	HYPOTHETICAL 33.3 KD P	1.67e+01
32	84	9.3	296	2	O69681	HYPOTHETICAL 72.6 KD P	1.67e+01
33	84	9.3	701	14	O65568	ANION EXCHANGER 2 A (F	1.67e+01
34	84	9.3	729	11	O60470	DIACYLGLYCEROL KINASE	1.67e+01
35	84	9.3	901	4	O60391	3' ORF.	2.17e+01
36	84	9.3	1117	4	O00542	NEUROMODULIN (FRAGMENT).	2.17e+01
37	83	9.2	147	11	O61639	MYOMODULIN PRECURSOR (	2.17e+01
38	83	9.2	248	14	O12396	LECITHIN:CHOLESTEROL A	2.17e+01
39	83	9.2	329	5	O27916	MALTASE-LIKE PROTEIN A	2.17e+01
40	83	9.2	370	5	O07974	PCBR.	2.17e+01
41	83	9.2	440	11	O35849	MHC CLASS I HEAVY CHAI	2.82e+01
42	83	9.2	498	5	O17021	VERY LARGE TEGUMENT PR	2.82e+01
43	83	9.2	551	2	P72405		
44	82	9.1	147	7	O46723		
45	82	9.1	3122	14	P89459		

#### ALIGNMENTS

RESULT 1  
ID O55241 PRELIMINARY; PRT; 130 AA.  
AC O55241;  
DT 01-JUN-1998 (TREMREL. 06, CREATED)  
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
DE HYPOCRETIN (PREPRO-OREXIN).  
GN HCRT.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98150861.  
RA SAKURAI T., AMEMIYA A., ISHII M., MATSUZAKI I., CHEMELI R.M.,  
RA TANAKA H., WILLIAMS S.C., RICHARDSON J.A., KOZLOWSKI G.P., WILSON S.,  
RA ARCH J.R.S., BUCKINGHAM R.E., HAYNES A.C., CARR S.A., ANNAN R.S.,  
RA MCNULTY D.E., LIU W.-S., TERRETT J.A., ELISHOURBAGY N.A., BERGMA D.J.,  
RA IYANAGISAWA M.;  
RT "Orexins and orexin receptors: a family of hypothalamic neuropeptides  
and G protein-coupled receptors that regulate feeding behavior.";  
RL CELL 92:573-585(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDOFF T.S., PEYRON C., GAO X.-B., FOYE P.E.,  
RA DANIELSON P.E., FUKUHARA C., BATTENBERG E.L.F., GAUTVIK V.T.,  
RA BARTLETT F.S. III, FRANKEL F.S., VAN DEN POL A.N., BLOOM F.E.,  
RA GAUTVIK K.M., SUTCLIFFE J.G.;  
RT "The hypocretins: hypothalamus-specific peptides with neuroexcitatory  
activity";  
RL PROC. NATL. ACAD. SCI. U.S.A. 95:322-327(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6J;  
RX MEDLINE; 98081872.  
RA DE LECEA L., KILDO

Matches 99; Conservative 13; Mismatches 11; Indels 1; Gaps 1;

Db 8 VSWATVTLTLLTLLPPAVLSPGAAQPLPDCCRQKTCSCRLYELLHGAGNHAAGIITLG 67  
 QY 1 VFWAAVTLTLLTLLTLLPPALLSGLVGDAPLPPDCCRQKTCSCRLYELLHGAGNHAAGIITLG 59  
 Db 68 KRPGPPGLOGRLQRLQASGNHAAAGILTMGRAGAPAPRLCPGRRCLAAASSVAPGG 127  
 QY 60 KRPGPPGLOGRLQRLQASGNHAAAGILTMGRAGALEPHEPCSGRCPCPTVTITAPRG 119  
 Db 128 RSGI 131  
 QY 120 GSGV 123

RESULT 5  
 ID O60609 PRELIMINARY; PRT; 400 AA.  
 AC O60609;  
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)  
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)  
 DE GDNF FAMILY RECEPTOR ALPHA 3.  
 GN GFRA3.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;  
 OC CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BALOH R.H., GORODINSKY A., GOLDEN J.P., TANSEY M.G., KECK C.L.,  
 RA POCESCU N.C., JOHNSON E.M. JR., MILBRANDT J.;  
 RL PROC. NATL. ACAD. SCI. U.S.A. 0:0-0(1998).  
 DR EMBL; AF051767; G2961632; -;  
 SQ SEQUENCE 400 AA; 44538 MW; 6DFB5381 CRC32;

Query Match 12.6%; Score 113; DB 4; Length 400;  
 Best Local Similarity 47.4%; Pred. No. 3.75e-03;  
 Matches 18; Conservative 9; Mismatches 9; Indels 2; Gaps 2;

Db 9 PLPPVVMILLTLLPPSPPLPAA-GDPLPTESLMN-SC 44  
 QY 2 PWAATVTLTLLTLLPPALLSGLVGDAPLPPDCCRQKTCSC 39

RESULT 6  
 ID O23352 PRELIMINARY; PRT; 679 AA.  
 AC O23352;  
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE HYPOTHETICAL 74.6 KD PROTEIN.  
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).  
 OC EUKARYOTA; VIRIDIPANTAE; SPRETOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;  
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;  
 OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;  
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 98121113.  
 RA BEVAN M., BANCROFT I., BENT E., LOVE K., GOODMAN H., DEAN C.,  
 RA BERKAMP R., DIRKSE W., VAN STAVEREN M., STIEKEMA W., DROST L.,  
 RA RIDLEY P., HUDSON S.A., PATEL K., MURPHY G., PIFANELLI P., WEDLER H.,  
 RA WEDLER E., WAMBUTT R., WEITZENEGGER T., POHL T.M., TERRY N.,  
 RA GIELEN J., VILLARROEL R., DE CLERCQ R., VAN MONTAGU M., LECHARNY A.,  
 RA AUBORG S., GY I., KREIS M., LAO N., KAVANAGH T., HEMPEL S., KOTTER P.,  
 RA ENTIAN K.D., RIEGER M., SCHAEFFER M., FUNK B., MUELLER-AUER S.,  
 RA SILVEY M., JAMES R., MONTFORT A., PONS A., PUIGDOMENECH P., DOUKA A.,  
 RA VOUKELATOU E., MILIONI D., HATZOPOULOS P., PIRAVANDI E., OBERMAIER B.,  
 RA HILBERT H., DUESTERHOFF A., MOORES T., JONES J.D.G., ENEVA T.,  
 RA PALME K., BENES V., RECHMAN S., ANSORGE W., COOKE R., BERGER C.,  
 RA DELSENY M., VOET M., VOLCKAERT G., MEWES H.W., KLOSTERMAN S.,

RA SCHUELLER C., CHALWATZIS N.;  
 RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of  
 RT Arabidopsis thaliana";  
 RL NATURE 391:485-488(1998).  
 DR EMBL; Z97337; E326841; -;  
 KW HYPOTHETICAL PROTEIN.  
 SQ SEQUENCE 679 AA; 74635 MW; B301B713 CRC32;

Query Match 11.8%; Score 106; DB 10; Length 679;  
 Best Local Similarity 32.3%; Pred. No. 3.22e-02;  
 Matches 20; Conservative 16; Mismatches 22; Indels 4; Gaps 4;

Db 598 KYCRSK-YETIHGQNDHNAADVLELAIKREMPAELL-R-ASLRHTNEDORNFILNVRSA 654  
 QY 35 KTCSCRLYELLHGAG-NHAAAGILTLGKRPPGPGLOGRLQRLQASGNHAAAGIITMGRRA 93  
 Db 655 SP 656  
 QY 94 GA 95

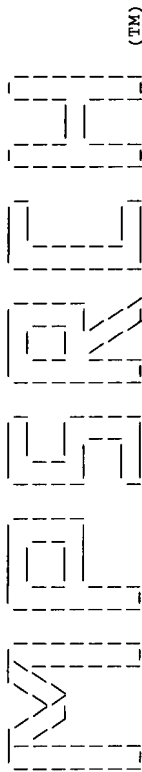
RESULT 7  
 ID O08463 PRELIMINARY; PRT; 641 AA.  
 AC O08463;  
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)  
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
 DE FRIZZLED PROTEIN HOMOLOG 1 (FZ-1).  
 GN FZ-1.  
 OS RATTUS NORVEGICUS (RAT).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;  
 OC SCIURIGNATHI; MURIDAE; MURINAE; RATTUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-OSTEOSARCOMA;  
 RX MEDLINE; 93094228.  
 RA CHAN S.D.H., KARP D.B., FOWLKES M.E., HOOKS M., BRADLEY M.S.,  
 RA VUONG V., BAMBINO T., LIU M.Y.C., ARNAUD C.D., STREWLER G.J.,  
 RA NISSENSON R.A.;  
 RT "Two homologs of the Drosophila polarity gene frizzled (fz) are  
 RT widely expressed in mammalian tissues";  
 RL J. BIOL. CHEM. 267:25202-25207(1992).  
 CC -1- FUNCTION: MAY BE INVOLVED IN TRANSDUCTION AND INTERCELLULAR  
 CC TRANSMISSION OF POLARITY INFORMATION DURING TISSUE MORPHOGENESIS  
 CC AND/OR IN DIFFERENTIATED TISSUES.  
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN KIDNEY, LIVER, UTERUS, OVARY  
 CC AND HEART. LOWER LEVELS SEEN IN BRAIN AND INTESTINE.  
 CC EXTREMELY LOW IN CALVARIA, MAMMARY GLANDS AND TESTIS.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED PREDOMINANTLY IN NEONATAL TISSUES,  
 CC AT LOWER LEVELS IN ADULT.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 DR EMBL; L02529; G310117; -;  
 KW DEVELOPMENTAL PROTEIN; TRANSMEMBRANE; GLYCOPROTEIN.  
 RN DOMAIN 72 312  
 FT TRANSMEM 313 338 POTENTIAL.  
 FT TRANSMEM 346 368 POTENTIAL.  
 FT TRANSMEM 397 425 POTENTIAL.  
 FT TRANSMEM 440 459 POTENTIAL.  
 FT TRANSMEM 483 508 POTENTIAL.  
 FT TRANSMEM 531 551 POTENTIAL.  
 FT TRANSMEM 594 616 POTENTIAL.  
 FT CARBOHYD 125 125 POTENTIAL.  
 FT CARBOHYD 225 225 POTENTIAL.  
 SQ SEQUENCE 641 AA; 71027 MW; EA083C54 CRC32;

Query Match 10.9%; Score 98; DB 11; Length 641;  
 Best Local Similarity 64.0%; Pred. No. 3.44e-01;  
 Matches 16; Conservative 2; Mismatches 6; Indels 1; Gaps 1;

Db 50 WARGLLLLLWLEAPLL-LGVRAQP 73  
 QY 3 WAAVTLTLLTLLPALLSGLVGDAP 27

[illegible]

\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 20 21:10:43 1999; MasPar time 4.50 Seconds  
Tabular output not generated. 249.097 Million cell updates/sec

Title: >US-08-938-548B-9  
Description: (1-28) from US08938548B.pap  
Perfect Score: 201  
Sequence: 1 RPQPPGLQGRLLQRLQANGNHAAGILTM 28

Scoring table: PAM 150  
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir60  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 30.698; Variance 52.943; scale 0.580

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query	ID	Description	Pred. No.
1	71	35.3	145	2	T00987	1.77e+00
2	71	35.3	187	2	B41654	1.77e+00
3	71	35.3	187	2	A41554	1.77e+00
4	70	34.8	511	2	T02269	2.52e+00
5	68	33.8	736	2	S47044	5.02e+00
6	68	33.8	879	2	H64888	5.02e+00
7	68	33.8	1015	2	T00730	5.02e+00
8	68	33.8	2115	2	S23647	5.02e+00
9	67	33.3	178	2	E69913	7.06e+00
10	67	33.3	442	2	S11712	7.06e+00
11	67	33.3	510	2	S41307	7.06e+00
12	67	33.3	525	2	JN0443	7.06e+00
13	67	33.3	528	2	JN0445	7.06e+00
14	67	33.3	1157	2	A55152	7.06e+00
15	66	32.8	150	2	D69081	9.90e+00
16	66	32.8	253	2	S49183	9.90e+00
17	66	32.8	319	1	S19248	9.90e+00
18	65	32.3	1420	2	T02644	1.38e+01
19	65	32.3	1692	2	A33988	1.38e+01
20	65	32.3	2509	2	G01880	1.92e+01
21	64	31.8	239	2	S25618	1.92e+01
22	64	31.8	256	1	PRH03	1.92e+01
23	64	31.8	263	2	S51155	1.92e+01

24	64	31.8	439	2	S61858	hrpE protein - Pseudo	1.92e+01
25	64	31.8	535	2	A46101	protein-tyrosine-phos	1.92e+01
26	64	31.8	548	2	B46101	protein-tyrosine-phos	1.92e+01
27	64	31.8	624	1	RDYCS7	sulfite reductase (fe	1.92e+01
28	64	31.8	1348	2	S27812	probable epidermal ce	1.92e+01
29	64	31.8	1348	2	A43917	probable epidermal ce	1.92e+01
30	64	31.8	1477	2	S64616	YOR1 protein - yeast	1.92e+01
31	63	31.3	255	2	S12255	hypothetical protein	2.67e+01
32	63	31.3	394	2	A42115	ribosomal protein S2,	2.67e+01
33	63	31.3	556	2	B64939	hypothetical protein	2.67e+01
34	63	31.3	702	2	C64835	hypothetical protein	2.67e+01
35	63	31.3	702	2	S77523	hypothetical protein	2.67e+01
36	63	31.3	812	2	S31521	collagen COL1A1-fres	2.67e+01
37	63	31.3	878	2	A41055	ecdysone receptor - f	2.67e+01
38	62	30.8	380	2	A42832	factor VIIa-associate	3.68e+01
39	62	30.8	466	2	S61292	transcription initiat	3.68e+01
40	62	30.8	652	2	I40676	transcription initiat	3.68e+01
41	62	30.8	684	2	A53019	collagen alpha 1(XVII	3.68e+01
42	62	30.8	767	2	T00360	hypothetical protein	3.68e+01
43	62	30.8	1288	3	JE0363	mitogen-activated pro	3.68e+01
44	62	30.8	1690	1	CGHUIB	collagen alpha 4(IV)	3.68e+01
45	62	30.8	1763	2	S16366	collagen alpha 2(IV)	3.68e+01

ALIGNMENTS

RESULT 1  
ENTRY #type complete  
TITLE T00987  
ORGANISM #formal\_name Arabidopsis thaliana #common\_name mouse-ear cress  
DATE 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 12-Feb-1999  
ACCESSIONS T00987  
REFERENCE Z14153  
#authors Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.  
#submision submitted to the EMBL Data Library, April 1998  
#description Arabidopsis thaliana chromosome II BAC T9J22 genomic sequence.  
#accession T00987  
#status preliminary; translated from GB/EMBL/DBJ  
#molecule\_type DNA  
#residues 1-145 #label ROU  
#cross-references EMBL:AC002505; NID:g27393379; PID:g27393379

GENETICS  
#map\_position II  
#introns 17/1; 49/3; 78/2; 123/2  
#note T9J22.21  
SUMMARY  
#length 145 #molecular-weight 15355 #checksum 3045  
Query Match 35.3%; Score 71; DB 2; Length 145;  
Best Local Similarity 39.1%; Pred. No. 1.77e+00;  
Matches 9; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

Db 116 PEKPGLEGWESIMEALGAHADS 138  
| | | | | : : : : : | | | :  
Qy 2 PGPPGLQGRLLQRLQANGNHAAG 24

RESULT 2  
ENTRY #type complete  
TITLE B41654  
ORGANISM #formal\_name Haemophilus parainfluenzae  
DATE 12-Jun-1992 #sequence\_revision 12-Jun-1992 #text\_change 03-Mar-1999  
ACCESSIONS B41654  
REFERENCE A41654  
#authors Kroll, J.S.; Langford, P.R.; Loynds, B.M.  
#journal J. Bacteriol. (1991) 173:7449-7457

```

6
ENTRY      H64888      #type complete
TITLE      membrane protein ydbH - Escherichia coli
ORGANISM   #formal_name Escherichia coli
DATE       12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
13-Nov-1998
ACCESSIONS H64888
REFERENCE   A64720
#authors   Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
            Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
            Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
            Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
            Y.
#journal   Science (1997) 277:1453-1462
#title     The complete genome sequence of Escherichia coli K-12.
#cross-references MIMD:97426617
#accession H64888
#molecule_type DNA
#status    nucleic acid sequence not shown; translation not shown
#residues  1-879 #label BLAT
#cross-references GB:AE000235; GB:U00096; NID:g1787643; PID:g1787646;
            UWGP:B1381
#experimental_source strain K-12, substrain MG1655
GENETICS   ydbH
FEATURE    8-24
SUMMARY    #domain transmembrane #status predicted #label TM01
            #length 879 #molecular-weight 96834 #checksum 5911

Query Match      33.8%; Score 68; DB 2; Length 879;
Best Local Similarity 45.8%; Pred. No. 5.02e+00;
Matches 11; Conservative 6; Mismatches 6; Indels 1; Gaps 1;

Db 396 GVDGRLOAIQAHENELGDFVLHM 419
|:|||||:|||||:|:|:|
QY 6 GLOGRLORLQANHGAAG-ILTM 28
|:|||||:|||||:|:|:|

7
ENTRY      T00730      #type complete
TITLE      hypothetical protein F22013.25 - Arabidopsis thaliana
ORGANISM   #formal_name Arabidopsis thaliana #common_name mouse-ear
            cress
DATE       12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
12-Feb-1999
ACCESSIONS T00730
REFERENCE   Z14200
#authors   Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.;
            Sun, H.; Conway, A.; Conway, A.; Kurtz, D.; Oji, O.; Shen,
            Y.K.; Toriumi, M.; Vysotskaia, V.; Yu, G.; Davis, R.W.;
            Federspiel, N.A.; Theologis, A.; Ecker, J.R.
#submission submitted to the EMBL Data Library, April 1998
#description Genomic sequence for Arabidopsis thaliana BAC F22013.
#accession  T00730
#status    preliminary; translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues  1-1015 #label SHI
#cross-references EMBL:AC003981; NID:g3063438; PID:g3063463
GENETICS
#map_position I
#introns
SUMMARY    #length 1015 #molecular-weight 111751 #checksum 6299

Query Match      33.8%; Score 68; DB 2; Length 1015;
Best Local Similarity 50.0%; Pred. No. 5.02e+00;
Matches 13; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Db 499 PTRGLRSSLRKRCNGNPTAATILT 524
|:|||||:|:|:|
QY 2 PGPPGLOGLRQLRLLQANGNHAAGILT 27
|:|||||:|:|:|

8
ENTRY      E69913      #type complete
TITLE      hypothetical protein yncC - Bacillus subtilis
ORGANISM   #formal_name Bacillus subtilis
DATE       05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
24-Sep-1998
ACCESSIONS E69913
REFERENCE   A69580
#authors   Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
            Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
            Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
            A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
            Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
            Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
            Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
            Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
            Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
            M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
            S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
            Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
            C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
            Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
            Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
            Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.;
            Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
            Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
            Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
            M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
            M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro,
            V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
            A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
            Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
            Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
            Seiwisch, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
            Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
            B.; Sorokin, A.; Taconi, E.; Takagi, T.; Takahashi, H.;
            Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
            Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
            Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
            Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
            Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
            K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;

```

Search completed: Fri Aug 20 21:10:58 1999  
Job time : 15 secs.